Differential Psychology

REVISED EDITION

DIFFERENTIAL PSYCHOLOGY

Individual and Group Differences in Behavior

BY *Anne Anastasi*
PROFESSOR OF PSYCHOLOGY
FORDHAM UNIVERSITY

AND *John P. Foley, Jr.*
VICE-PRESIDENT, THE PSYCHOLOGICAL CORPORATION
AND DIRECTOR, INDUSTRIAL DIVISION

THE MACMILLAN COMPANY : NEW YORK

Preface to the Revised Edition

IN KEEPING WITH the rapid growth of differential psychology during the past decade, the present edition represents a thorough revision and a considerable enlargement of the original book. Four new chapters have been added, covering basic concepts of psychological testing (Ch. 2), biological and psychological factors in simple behavior development (Ch. 5, 6) and the effects of schooling upon intelligence (Ch. 8). The inclusion of the new Chapters 5 and 6 reflects a greater emphasis upon the developmental approach in the study of behavior differences, an approach also illustrated by the increasing number of longitudinal studies reported throughout the book. Much more material is now available on individual and group differences in personality characteristics. These findings have been discussed, together with the results on intellectual differences, in the appropriate sections. The content of the chapters on trait organization (Ch. 15) and socio-economic differences (Ch. 23) is substantially new, most of the studies covered having been conducted during the last ten years. The remaining chapters have likewise been reorganized and rewritten in order to incorporate and integrate recent developments in each area. The present edition has also drawn more extensively upon recent findings in genetics, anthropology, and sociology.

Partly as a result of protracted controversies in certain areas, the past decade has witnessed an increasing methodological and theoretical sophistication among workers in differential psychology. More rigid standards of experimental control have been demanded, traditional procedures have been scrutinized and challenged, and a greater concern for sharply defined concepts has been evidenced. In recognition of these developments, the present book places more emphasis upon methodological problems and basic concepts. This is especially illustrated in the reorganized Part I, whose aim is to give the student a brief over-all introduction to certain important concepts of psychological testing, heredity and environment, and the nature of individual differences, prior to the presentation of specific data in Parts II

and III. Even in the subsequent treatment of individual or group differences, however, the obtained findings are always interpreted operationally in the light of experimental conditions and methodology.

While keeping abreast of the many changes in the field, the present book retains the fundamental objectives of the original edition. First, differential psychology is presented, not as a separate field of psychology, but as one approach to the *understanding of behavior.* Its fundamental questions are no different from those of general psychology. It is apparent that if we can explain why individuals react differently from one another, we shall understand why each individual reacts as he does. The data of differential psychology should thus help to clarify the basic mechanisms of behavior. It is primarily from this point of view that the problems of individual and group differences are surveyed in the present text.

A second aim of the book has been to *coordinate* the various topics which have usually been joined together loosely under the caption of "individual differences." The phenomenal development of differential psychology during the past decade has resulted in an increasing specialization of interest among research workers and a frequent disregard of the broader implications of the data. The mutual interrelation of the various problems has often been obscured by the accumulation of data at a more rapid pace than they could comfortably be assimilated. For this reason, the writers have endeavored to bear constantly in mind the interrelationships among different types of investigations and have attempted to present a systematic organization and integration of the material. No chapter stands alone. Each is related to what preceded it and to what is to follow.

Thirdly, it has been our aim to report the major findings of differential psychology in a form *readily comprehensible to the college student.* Our purpose has been to present the material clearly and interestingly, at the same time avoiding the errors of significant omission and falsification which so frequently characterize many attempts at popularization. Since much of the knowledge in this field is found only in highly technical sources, certain topics have customarily received cursory or superficial treatment in texts on individual differences. Even the more advanced student of psychology who has specialized in other phases of the subject often finds it impossible to keep informed on current developments in certain branches of differential psychology. The writers are convinced, however, that a non-technical and easily

comprehensible presentation of such topics is both feasible and desirable. An understanding of the basic concepts and major findings within any area need not be limited to those who have mastered its specialized techniques.

One further point deserves special mention. The present book is not intended to be a literature survey. First and foremost. it is a *textbook* designed to develop in the student the intellectual skills needed for understanding and evaluating the data of differential psychology. Throughout the book, special emphasis has been placed upon the examination of common pitfalls and sources of error in the interpretation of obtained results. We have thus hoped to provide the student with certain *tools whereby he may evaluate for himself a set of data with which he is confronted.* This would seem to be far more important than the mere presentation of a body of facts. The development of critical ability and of a dispassionate and objective attitude toward human behavior is more urgently needed today than ever before.

The writers are pleased to acknowledge the cooperation of several colleagues in the preparation of this revision. Professors Robert T. Rock, Jr., and Dorothea McCarthy of the Department of Psychology, Fordham University, contributed many valuable suggestions and provided assistance in countless other ways. The writers are indebted to Professor Robert L. Thorndike of Teachers College, Columbia University, for his intensive reading of the chapter on *Schooling and Intelligence* and for his constructive comments in this area. Thanks are extended to Professor Charles A. Berger, Chairman of the Department of Biology, Fordham University, for his critical reading of the sections dealing with genetics. The writers wish to express their appreciation to Professor Frank Lee of the School of Engineering, Columbia University, for his skillful and painstaking preparation of the illustrations. Finally, grateful acknowledgment is made to Mrs. Enrica Tunnell, Librarian, Psychology Reading Room, Columbia University, for her ready aid in many bibliographic matters.

A. A.
J. P. F.

Table of Contents

PART II. ANALYSIS OF INDIVIDUAL DIFFERENCES

PART I HISTORICAL AND METHODOLOGICAL ORIENTATION

CHAPTER

1

Historical Introduction

MAN HAS ALWAYS BEEN AWARE of differences among his fellow beings. He has, to be sure, entertained various theories, beliefs, or superstitions regarding the causes of such differences, and has interpreted them differently according to his own traditional background, but he has at all times accepted the fact of their existence. Among primitive peoples, unusual deviations in behavior are clearly recognized. Thus many primitive groups acknowledge exceptional artistic talent among their members and encourage the development of specialized artists. The presence of hysterical or epileptoid symptoms, paranoid trends, and similar peculiarities of behavior has frequently been regarded as an index of religious or magical powers and has been treated accordingly. At any level of cultural development, specialization of labor itself implies a tacit assumption of differences among people.

Nor is this response to individual differences limited to the human species. Instances from infrahuman behavior can readily be found. The acceptance of certain individuals as "leaders" by herds of elephants, buffaloes, and similar gregarious animals has been widely discussed in the literature both of fact and of fiction. In communities of baboons, a certain member is posted as "sentinel" to watch for the approach of danger and warn the others by conventional cries. The frequently described "hacking" or hen-pecking behavior of chickens is another case in point. A definite relationship of social domination is often displayed by chickens in the barnyard, this fighting or "hacking" behavior usually centering about the acquisition of food. In such cases, A will attack B, although the reverse will not occur. Violent conflicts often ensue when the authority of the chief "hacker" in the group is disputed. These and many other examples illustrate the prevalence of differential responses to individuals within one's own group.

The objective and quantitative investigation of individual differences in behavior phenomena is the domain of differential psychology. What

is the nature and extent of such differences? What can be discovered as to their causes? How are the differences affected by training, growth, physical conditions? In what manner are the differences in various traits related to one another, or organized? These are some of the fundamental questions raised by differential psychology.

Part I of the present book furnishes a brief *historical and methodological orientation* to the field of differential psychology, including an introduction to its fundamental *concepts*. Among the concepts treated are those underlying psychological testing, the distribution of individual differences, and heredity and environment. The analysis of *specific data on individual differences* forms the content of Part II, this analysis being developed in terms of the effect of different factors upon behavior development. Thus the contributions of such factors as training, age, familial relationships, and structural correlates are considered in turn, followed by a discussion of the nature and interrelationships of psychological traits.

Differential psychology is also concerned with an analysis of the *nature and characteristics of major traditional groupings*, such as the subnormal and the genius, the sexes, and racial, national, and cultural divisions. This furnishes the subject matter of Part III. The study of such group differences serves a threefold purpose. In the first place, one cannot ignore the fact that such groupings are being made in the practical realm of everyday life. These distinctions cannot be swept away casually on the grounds that, perhaps, the study of individual differences reveals no need for them or for any sharp divisions into clear-cut categories. Certain groups are recognized and responded to as distinctive in our present-day society. For a purely practical reason, therefore, these groups must be investigated, in the hope that the specific findings may throw some light upon their nature and possibly further a more intelligent practical understanding of them.

Secondly, the comparative investigation of different groups should help to clarify the fundamental problems of individual differences in general. In such groups we can see the principles of individual differences in operation and can note their effects. Group differences in behavior, when considered in conjunction with other concomitant differences among the groups, furnish an excellent available means of analyzing the causes of variability.

Thirdly, the comparison of a psychological phenomenon as it occurs in different groups may contribute toward a clearer under-

standing of the phenomenon itself. The findings of general psychology, when tested in widely varying groups, are sometimes found to be not so general as was supposed. To study a phenomenon in all its varied manifestations is to have a better grasp of its essential nature.

• Notwithstanding the early and widespread recognition of individual differences in the practical adjustments of everyday life, the systematic investigation of such differences is a relatively recent development in psychology. We may therefore begin by considering the principal historical developments in the field of differential psychology.

INDIVIDUAL DIFFERENCES IN PRE-EXPERIMENTAL PSYCHOLOGICAL THEORY [1]

One of the earliest instances of the explicit recognition of individual differences is to be found in the *Republic* of Plato. A fundamental aim of Plato's ideal state is, in fact, the assignment of individuals to the special tasks for which they are suited. In Book II of the *Republic* appears the following statement: "Really, I said, it is not improbable; for I recollect myself, after your answer, that, in the first place, no two persons are born exactly alike, but each differs from each in natural endowments, one being suited for one occupation and another for another" (11, p. 60). Plato proposes a series of "actions to perform" for use as tests of military aptitude on those who are to be the soldiers of his ideal state. These actions are designed to sample the various traits considered essential to military prowess, and represent the first systematic description of an aptitude test on record.

Nor did the versatile genius of Aristotle overlook individual variation. He discusses at some length group differences, including species, racial, social, and sex differences in mental and moral traits. In many of his works there is also an implicit assumption of individual differences, although it is interesting to note that Aristotle does not offer any extensive treatment of these differences as such. One gets the impression that he regards the existence of individual variation as too

[1] To supplement the brief historical sketch of the study of individual differences given in the present and following sections, the reader is referred to any of the standard works on the history of psychology, such as Boring (5), Murphy (31), and Rand (37).

The numbers in parentheses here and throughout the book refer to the numbered References at the end of the respective chapter.

obvious to need special mention. That he attributes such differences at least in part to innate factors seems to be indicated by a number of statements such as the following:

> Perhaps, then, some one may say, "Since it is in my power to be just and good, if I wish I shall be the best of all men." This, of course, is not possible. . . . For he who wills to be best will not be so, unless Nature also be presupposed (38, *Magna Moralia,* 1187b).

Throughout the several *Ethics* of Aristotle, there appear passages which imply individual variation. The following statement, for example, leaves little doubt regarding Aristotle's position on this point:

> After these distinctions we must notice that in everything continuous and divisible there is excess, deficiency, and the mean, and these in relation to one another or in relation to us, e.g., in the gymnastic or medical arts, and in those of building and navigation, and in any sort of action, alike scientific and non-scientific, skilled and unskilled. For motion is continuous, and action is motion (38, *Ethica Eudemia,* 1220b).

Aristotle then proceeds to describe the characteristics of men possessing an excess or a deficient amount of various traits such as irascibility, audacity, shamelessness, and others.

In the Scholasticism of the Middle Ages, individual differences were largely neglected. Philosophical generalizations regarding the nature of the mind were formulated through "rational," or speculative, rather than "empirical" means. The observation of individuals thus played little or no part in the development of such doctrines. Of particular interest for differential psychology is the "faculty psychology" advanced by St. Augustine and Thomas Aquinas. Such "faculties" as "memory," "imagination," and "will" have been regarded by some as the precursors of the traits and factors currently identified through statistical analysis of test scores. These empirically determined factors, however, differ in a number of important ways from the rationally derived faculties of Scholastic philosophy (cf. Ch. 15).

The many varieties of Associationism which flourished from the seventeenth to the nineteenth centuries likewise took little heed of individual differences. It was with the elaborate mechanics whereby "ideas" become associated, giving rise to complex mental processes, that the associationists were primarily concerned. Their statements were general principles with no allowance for individual variation. Bain, the last of the so-called pure associationists, does, however, give

some attention to individual differences in his writings. The following passage is taken from his book on *The Senses and the Intellect* (1855): "There is a natural force of adhesiveness, specific to each constitution, and distinguishing one individual from another. This property, like almost every other assignable property of human nature, I consider to be unequally distributed" (2, p. 237).

A simultaneous development in educational theory should probably be included at this point. In the writings and practices of a group of "naturalist" educators of the late eighteenth and early nineteenth centuries, including Rousseau, Pestalozzi, Herbart, and Froebel, there is found a definite shift of interest to the individual child (cf. 29). Educational policies and methods were to be determined, not by external criteria, but by direct observation of the child and his capacities. The emphasis still seemed to be, however, on the observation of the individual as representative of individuals in general, rather than as distinct from other individuals. Although statements can be found in the writings of these educators to the effect that individuals differ and that their education should be adapted to these differences, still the emphasis is laid more heavily upon free, "natural" education in contrast to externally and arbitrarily imposed procedures, rather than upon individual differences themselves. The term "individual" is often used to mean simply "human nature."

Finally, mention may be made of the various treatises on race and racial psychology which appeared in the late eighteenth and early nineteenth centuries. Discussions of race differences are to be found in the works of such writers as Buffon, Herder, and de Gobineau, the last having been especially influential in determining subsequent popular beliefs about race.

THE PERSONAL EQUATION IN ASTRONOMY

Curiously enough, the first systematic measurements of individual differences were undertaken not in psychology but in the old and time-honored science of astronomy. In 1796, Maskelyne, the astronomer royal at the Greenwich Observatory, dismissed Kinnebrook, his assistant, because the latter observed the times of stellar transits nearly a second later than he did. The method employed at the time to make such observations was the "eye and ear" method. This method involved not only coordination of visual and auditory impressions, but

also rather complex spatial judgments. The observer noted the time to a second on the clock, then began to count seconds with the heard beats of the clock, at the same time watching the star as it crossed the field of the telescope. He noted the position of the star at the last beat of the clock just before it reached the "critical" line in the field; then, similarly, he noted its position with the first beat immediately after it had crossed that line. From these observations, an estimate was made in tenths of a second of the exact time when the star crossed the critical line. This was the accepted procedure and was regarded as accurate to one- or two-tenths of a second.

In 1816 Bessel, astronomer at Königsberg, read of the Kinnebrook incident in a history of the Greenwich Astronomical Observatory, and became interested in measuring what came to be known as the "personal equation" of different observers. Originally, the personal equation referred to the difference in seconds between the estimates of two observers. Bessel collected and published data on several trained observers, and pointed out not only the presence of such a personal equation or error when comparing any two observers, but also the variability in the equation from time to time. This represents the first published record of quantitative data on individual differences.

Many astronomers followed up Bessel's measurements. In the latter half of the nineteenth century, with the introduction of chronographs and chronoscopes, it became possible to measure the personal equation of a given observer without reference to any other observer. The attempt was made to reduce all observations to their objectively correct values without reference to a system of time based upon one observer as a standard. Astronomers also undertook an analysis of the various conditions which affected the size of the personal equation. It was this latter problem, rather than the measurement of individual differences, which was taken up by the early experimental psychologists in their studies of "reaction time."

THE RISE OF EXPERIMENTAL PSYCHOLOGY

During the latter half of the nineteenth century, psychology began to venture away from its armchair and enter the laboratory. Most of the early experimental psychologists were physiologists whose experiments gradually came to take on a psychological tinge. As a result, both the viewpoints and the methods of physiology were frequently carried

over directly into the infant science of psychology. In 1879, Wilhelm Wundt established the first laboratory of experimental psychology at Leipzig. Experiments of a psychological nature had been performed previously by Weber, Fechner, Helmholtz, and others, but Wundt's laboratory was the first to be devoted exclusively to psychology and to offer facilities for training students in the methods of the new science. Students from many parts of the world were attracted to Wundt's laboratory, and upon their return founded laboratories in their own countries.

The problems investigated in these early laboratories testify to the close kinship of experimental psychology with physiology. The study of visual and auditory sensation, reaction time, psychophysics, and association constituted nearly the entire field of experimentation. It was characteristic of the early experimental psychologists either to ignore individual differences or to regard them simply in the nature of experimental "errors." The greater the individual variation in a phenomenon, the less accurate would be the generalizations regarding such a phenomenon. The extent of individual differences would thus represent the "probable error" to be expected in the application of the general formulations.

It is apparent that the rise of experimental psychology shifted the emphasis away from the study of individual differences rather than toward it. Its one contribution to the development of a differential psychology is to be found in its demonstration that psychological phenomena are amenable to objective and even quantitative investigation, that psychological theories can be tested by actual data, that psychology, in short, could become an empirical science. Such a step was required before theories about the individual could be replaced by studies on individual differences.

GALTON AND THE BIOLOGICAL INFLUENCE

With the spread of Darwinism in the late nineteenth century, psychology became increasingly biological in its approach. One of the most widely known of Darwin's followers was Sir Francis Galton, who first attempted to apply the evolutionary principles of variation, selection, and adaptation to the study of human individuals. Galton's scientific pursuits were many and varied, but they were unified by an underlying interest in the study of heredity. The science of eugenics, whose

aim is the control and direction of human evolution, was originated by Galton. In 1869, he published a book entitled *Hereditary Genius,* in which, by the application of the now well-known family history method, he tried to demonstrate the inheritance of specific talents in various fields of work. In connection with the study of human inheritance, it soon became apparent that related and unrelated individuals must be measured, objectively and in large numbers, in order to discover the degrees of resemblance among them For this purpose, Galton devised numerous tests and measure and in 1882 established his famous anthropometric laboratory at South Kensington Museum in London There, for the payment of a small fee, individuals could be tested in sensory discrimination, motor capacities, and other simple processes

Through the measurement of sensory processes, Galton hoped to arrive at an estimate of the subject's intellectual level In the *Inquiries into Human Faculty* a collection of miscellaneous essays published in 1883, he wrote "The only information that reaches us concerning outward events appears to pass through the avenue of our senses, and the more perceptive the senses are of difference, the larger is the field upon which our judgment and intelligence can act" (13, p 27) And again on the basis of findings on the inferior sensitivity of idiots, he observes that sensory discriminative capacity "would on the whole be highest among the intellectually ablest" (13, p 29) For this reason, measures of sensory capacity, such as vision and hearing, constituted a relatively large portion of the tests which Galton constructed and employed Among these tests may be mentioned the Galton bar for visual discrimination of length, the Galton whistle for the determination of the highest audible pitch, kinaesthetic discrimination tests based on the arrangement of a series of weights as well as tests of strength of movement, speed of simple reactions and many others of a similar nature

Galton also initiated the use of "free association" tests, a technique which was subsequently adopted and further developed by Wundt Galton's study of mental imagery is well known and represents the first extensive psychological use of questionnaire methods In this questionnaire, the subject was directed "to think of some definite object—suppose it is your breakfast table as you sat down to it this morning—and consider carefully the picture that rises before your mind's eye" (13, p 84) They were then to describe the picture with

reference to illumination, definition, and coloring. Wide individual and group differences were revealed by this analysis of imagery.

A further and very significant contribution of Galton to differential psychology was his development of statistical methods for the analysis of the data of individual differences. Formerly, statistics had been chiefly the tool of the trained mathematician and the professional gambler. Statistical techniques were not available in a form which would enable the mathematically untrained worker in the biological sciences to employ them. Galton realized the need for such techniques and developed many of the statistical procedures in current use today. This phase of his work has been extended and increased in scope by many eminent students, chief among whom is Karl Pearson, who succeeded Galton as director of the anthropometric laboratory in 1911.

EARLY EXPERIMENTATION WITH TESTS

The term "mental test" was first employed in 1890 by Cattell, in an article entitled *Mental Tests and Measurements* (9). James McKeen Cattell was an American student of Wundt. In 1888, having obtained his doctorate at Leipzig, he returned to this country, where he was instrumental both in the spread of experimental psychology and in the development of mental testing. He had also come under the influence of Galton's work in test construction and statistics. Thus in Cattell we find a convergence of two contemporary movements in psychology: the rise of the experimental method and the measurement of individual differences. It was characteristic of all the early American mental tests that they developed in the psychological laboratory and partook of the nature of the experimental psychology of the time. This was not true of many of the tests developed in other countries.

In addition to his experiments on reaction time, attention span, controlled association, reading, psychophysics, and similar problems, Cattell constructed a series of tests which were administered for many years to freshmen and seniors at Columbia College. This series included the following tests:[2] (1) strength of grip; (2) rate of arm movement; (3) two-point threshold on the back of the hand; (4) amount of pressure required to produce pain on the forehead; (5) least noticeable difference in weights, (6) reaction time to sound; (7) speed of color naming; (8) bisection of a 50-cm. line; (9) repro-

[2] For a fuller description, cf. Cattel and Farrand (10).

duction of a 10-second time interval; (10) auditory memory span for letters. This list is typical of the various test series which appeared at the time.

In 1891, Münsterberg (30) described a series of tests which he had employed on school children. Tests of reading, controlled association of various sorts, judgment, memory, and other simple mental processes were included. At the 1893 Columbian Exposition at Chicago, Jastrow administered a series of sensory, motor, and simple perceptual tests to all persons interested. Norms of physical growth and mental development were presented with the tests (cf. 34, 35).

What is probably the first attempt to evaluate test scores in terms of an independent criterion is to be found in the study by Bolton (4) reported in 1892. Bolton analyzed data collected by Boas on about 1500 school children. The children's memory spans were compared with their teachers' estimates of "intellectual acuteness," very little correspondence being found. Gilbert (16), in 1893, compared teachers' estimates of "general ability" on some 1200 children with their scores on eight tests of sensory and motor functions, reaction time, sensory memory, and suggestibility. Three years later, Gilbert (17) described some additional tests and reported the results obtained with them on several hundred children. The data were analyzed in respect to sex differences, intellectual growth, and the relationship of mental and physical development.

In Germany, Oehrn (33), a pupil of Kraepelin, published in 1889 the results of an intensive study of a series of tests on ten subjects. The tests had been rather arbitrarily selected to measure perception, memory, association, and motor functions. In 1895, Kraepelin (28) proposed a set of traits which he regarded as basic in the characterization of any individual. He also devised tests for the measurement of these traits, most of the tests involving simple arithmetic operations. These tests were of rather dubious validity for measuring the traits in question, and in addition they were quite impracticable, some of them requiring several days for their completion.

Research on mental tests was also being conducted simultaneously under the direction of the Italian psychologist Ferrari. In an article appearing in 1896, some of these tests were described (20). They included measures of vasomotor activity, motor strength and skill, range of apprehension, description of pictures, and temporal estima-

tion. Interesting individual differences were reported in many of these tests.

BEGINNINGS OF DIFFERENTIAL PSYCHOLOGY

At the turn of the century, differential psychology had begun to assume definite shape. In 1895 Binet and Henri published an article entitled *La psychologie individuelle* (3), which represents the first systematic analysis of the aims, scope, and methods of differential psychology. Their opening sentence suggests the status of this branch of psychology at the time. It reads: "We broach here a new subject, difficult and as yet very meagerly explored" (p. 411). Binet and Henri put forth as the two major problems of differential psychology, first, the study of the nature and extent of individual differences in psychological processes; and secondly, the discovery of the interrelationships of mental processes within the individual, so that we may arrive at a classification of traits and determine which are the more basic functions.

In 1900 appeared the first edition of Stern's book on differential psychology, under the title *Über Psychologie der individuellen Differenzen* (42). Part I deals with the nature, problems, and methods of differential psychology. Within the scope of this branch of psychology Stern included differences among individuals as well as among racial and cultural groups, occupational and social levels, and the two sexes. The fundamental problem of differential psychology he characterized as threefold: (1) What is the nature and extent of differences in the psychological life of individuals and groups? (2) What factors determine or affect these differences? In this connection he mentioned heredity, climate, social or cultural level, training, adaptation, etc. (3) How are the differences manifested? Can they be detected by such indices as handwriting, facial conformation, etc.? Stern also included a discussion of the concepts of psychological type, individuality, and normality and abnormality. Under the methods of differential psychology, he gave an evaluation of introspection, objective observation, the use of material from history and poetry, the study of culture, quantitative testing, and experiment. Part II contains a general discussion and some data on individual differences in various psychological traits, from simple sensory capacities to more complex mental processes and emotional characteristics. Stern's book appeared

in a highly revised and enlarged edition in 1911, and again in 1921, under the title of *Die Differentielle Psychologie in ihren methodischen Grundlagen* (43).

In America committees were being appointed to investigate testing methods and to sponsor the accumulation of data on individual differences. At its 1895 meeting, the American Psychological Association appointed a committee "to consider the feasibility of cooperation among the various psychological laboratories in the collection of mental and physical statistics" (10, p. 619). In the following year, the American Association for the Advancement of Science established a standing committee to organize an ethnographic survey of the white population in the United States. Cattell one of the members of this committee, pointed out the importance of including psychological tests in this survey and suggested that its work be coordinated with that proposed by the American Psychological Association (10, pp. 619–620).

The application of the newly devised mental tests to various groups was also getting under way. R. L. Kelly (25) in 1903 and Norsworthy (32) in 1906 compared normal and feebleminded children on sensori-motor and simple mental tests, and called attention to the continuous gradation in ability which exists between these groups, the feeble-minded not constituting a distinct category. In 1903 appeared Thompson's *The Mental Traits of Sex* (47), the result of several years' testing of men and women with a variety of tests. This represents the first comprehensive investigation on psychological sex differences. Tests of sensory acuity, motor capacities, and a few simple mental processes were also being administered for the first time to various racial groups. A few scattered investigations appeared before 1900. In 1904 Woodworth (50) and Bruner (7) tested several primitive groups at the St. Louis Exposition. In the same year appeared Spearman's original article putting forth his Two-Factor theory of mental organization and introducing a statistical technique for investigating the problem. Thus, shortly after the beginning of the century, the foundations had been laid for virtually every branch of differential psychology.

INTELLIGENCE TESTING

The intelligence test is a product of the twentieth century. The early mental tests were predominantly sensori-motor or very simple in

nature. This was no doubt a carry-over from the sensationism current in the psychological laboratories of the time. Complex mental processes were believed to be best understood by analyzing them into their elementary components, usually of a sensory nature. Most of the efforts of the early experimentalists were therefore devoted to the study of simple sensori-motor reactions, and this influence left its mark on the newly developing mental tests.

Binet and Henri, in their 1895 article (3), were the first definitely to point out the need for more complex tests to measure "intelligence." They examined the five most comprehensive current test series, those of Cattell, Münsterberg, Jastrow, Kraepelin, and Gilbert, and found all of them greatly overweighted with sensory tests and lacking in tests of complex processes. From an analysis of the available data, they concluded that individual differences are more marked in complex tasks and that the latter are therefore better suited to the study of such differences. Partly to remedy this deficiency in the current tests, Binet and Henri described ten types of tests which in their opinion would yield the largest and most significant individual differences. The series included tests of memory, mental imagery, imagination, attention, comprehension, suggestibility, æsthetic appreciation, moral feelings, muscular force and force of will, and motor ability and visual discrimination. The entire series, according to the authors, would require only from one to one and one-half hours.

In 1897, Ebbinghaus (12) proposed a theory to the effect that intelligence is the ability to combine or integrate the items of experience, and offered the sentence completion test as a technique for measuring this ability. In this test, the subject is presented with sentences in which certain of the words are missing and he is required to fill in the proper words. In experiments on German school children, Ebbinghaus had found this test more effective than simpler tests of calculation and memory. The completion test showed the most regular increase in score with age and it was also the only one of the tests employed which differentiated clearly among those pupils within each grade whose scholastic standing was good, average, or poor. Binet's contention for the superiority of the more complex tests in differential psychology was thus corroborated.

Two American studies of this period lent further support to Binet's statements. One of these studies (39) was conducted by Sharp, a student of Titchener, and was designed as a specific investigation of

the conclusions of Binet and Henri. A set of tests, modeled largely on those of Binet and Henri, was administered to seven advanced psychology students. The experiment was very intensive and included the repetition of similar tests on different days to determine the consistency of the processes tested. In general, although the need for further controls and refinements was suggested, the tests proved satisfactory and yielded sizable individual differences despite the homogeneous and select nature of the group. Sharp concluded: "We concur with Mm. Binet and Henri in believing that individual psychical differences should be sought for in the complex rather than in the elementary processes of mind. and that the test method is the most workable one that has yet been proposed for investigating these processes" (39, p. 390). A few years later, Wissler (49) published the results of his correlation analysis of the data collected in Cattell's laboratory. The correlations showed "little more than a chance relation" among the tests, and also a negligible correspondence with academic grades. Thus the inadequacy of the simple tests originally employed was further demonstrated.

Against this background of theory and data appeared the first intelligence scale. In 1904 the French Minister of Public Instruction appointed a committee to investigate the causes of retardation among public school children. Binet was one of the members of this committee. As a direct outgrowth of his work in this connection, Binet published, in collaboration with Simon, the 1905 scale for measuring intelligence. This scale consisted of 30 problems arranged in a rough order of difficulty. In 1908 appeared Binet's first revision of the scale, in which the tests were grouped into age levels and the concept of "mental age" [3] was introduced. The scale was again revised in 1911, the year of Binet's untimely death.

The Binet tests have been translated into most of the principal languages and their use has spread over every continent. In America, several different revisions have appeared, of which the most widely known is undoubtedly the Stanford-Binet, prepared by Terman and his associates at Stanford University. The intelligence quotient (IQ), found by dividing the child's mental age by his chronological age, was

[3] The child's score on an age scale is expressed as a mental age (MA). If, for example, he passes successfully all of the tests assigned to the 10-year level, he has a mental age of 10, regardless of what his chronological age may be.

first employed in 1916 in the Stanford-Binet, although the use of such a ratio had been previously discussed by Stern and others. The Stanford-Binet itself has undergone repeated revision, its latest revision (often referred to as the Terman-Merrill Scale) having appeared in 1937. This scale is available in two equivalent forms, L and M, which can be employed interchangeably. Its age range extends from two years to the adult mental level of 15. There are in addition three "Superior Adult" levels of increasing difficulty (cf. 46).

Of special interest is the 1922 Kuhlmann-Binet revision which extended the scale downward to the three months age level.[4] The construction of scales for measuring the intelligence of very young children represents one of the most recent developments in psychological testing.[5] Such tests have now become differentiated into *infant tests,* covering the period from birth to about 1½ years, and *preschool tests,* designed primarily for children between the ages of 1½ and 5. The former consist essentially of series of developmental norms which can be applied to an evaluation of the child's everyday behavior in such activities as crawling, walking, sitting up, standing, picking up and manipulating objects, recognizing colors and shapes, and acquiring the use of language. Among the most accurately established and extensive norms are those prepared by Gesell and his co-workers at Yale University (14, 15), where hundreds of infants have been periodically examined in practically every type of behavior under carefully controlled conditions. Preschool tests present the child with simple, standardized tasks, many of which involve motor coordination, the development of perceptual responses, and the understanding of spoken language. Among the best-known tests in current use for the preschool level are the Merrill-Palmer Scale (44) and the Minnesota Preschool Tests (19).

A relatively early development in the history of intelligence testing was the construction of *performance scales.* It was soon realized that the Binet type of test, depending so largely upon language, is not

[4] The 1939 revision of the Kuhlmann Tests of Mental Development represents an extensive restandardization of items from various sources, particularly the Kuhlmann-Binet and the Gesell normative scales.

[5] It is interesting to note, however, that as early as 1887 a series of developmental standards and simple tests for judging the mental level of infants during the first three years was worked out by an American physician, Dr. S. E. Chaille. The concept of mental age seems to have been implicit in his treatment of the data, although this term was not employed (cf. 18).

suitable for testing illiterates, the foreign-speaking, the deaf, or those who have speech disabilities. Performance tests were designed to meet this need, as well as to supplement the Binet type of test for a better-rounded picture of the individual's abilities. Among the earliest standardized series of performance tests was the Pintner-Paterson Scale (36), appearing in 1917. It consisted of 15 tests which could be administered without the use of either oral or written language on the part of either examiner or subject. Blocks, pictures of simple objects or scenes, and geometric forms are the principal materials of such tests. A more recently developed test is the Arthur Performance Scale, which originally consisted of a restandardization of ten of the Pintner-Paterson tests and of certain other tests taken from previously available series. The Arthur Scale has itself undergone considerable revision and modification during the past twenty years, the Revised Form II having been published in 1947 (1).

Mention may also be made of the recently developed Wechsler-Bellevue Scale (48), consisting of both verbal and performance tests and designed especially for the testing of *adults*. In content the test represents a combination of the Binet type and the performance type of item. Norms are given for ages 10 to 59. A special feature of this test is the suggested qualitative analysis of specific response patterns in terms of traditional clinical syndromes.

GROUP TESTING

All the intelligence tests discussed in the preceding section are "individual tests" in the sense that only one subject can be tested at a time. Furthermore, owing to the nature of these tests, a highly trained examiner is usually required to administer them. Testing on a large scale could not be conducted under these conditions. Data on such problems as sex and race differences, for example, which require the investigation of large samples, would be very slow in accumulating.

The advent of the group intelligence scale was probably the chief factor in the widespread popularization of mental testing. The group test is designed with a view to its general use. It is not only adapted to the simultaneous testing of large groups, but it is also relatively easy to administer and score. The impetus for the development of group tests was furnished in 1917 by the pressing need of testing over

one and one-half million men in the United States Army during World War I. A quick, rough classification in respect to intelligence was necessary for many purposes. Discharge because of serious mental defect, assignment to labor battalions requiring only low-grade work, admittance to officers training camps, and a number of similar problems required a knowledge of the intellectual level of the soldier.

Accordingly, a committee was appointed by the American Psychological Association to devise a test suited to this purpose. The committee consisted of five psychologists who were specialists in mental testing, and was under the direction of Robert M. Yerkes. All the available material on mental tests was examined for its suitability to the needs of the army testing program. An important source of such material was an unpublished group scale previously developed by Otis, which he made available to the government. The final outcome of the research of the army psychologists was the Army Alpha and the Army Beta. The former was the more widely used of the two; the latter is a non-language scale, and was designed for testing illiterates and foreigners unfamiliar with English.

After the close of World War I, new intelligence tests were constructed at a rapid rate. Soon special tests were available for elementary school children as well as kindergarten and preschool levels, high school and college students, and unselected adults. Mental testing attained undreamed-of proportions. School teachers were now considered to be qualified to administer the newly simplified tests. Large-scale school surveys were initiated; college freshmen were tested as part of the routine of admission, the general public became intelligence-test-conscious, and the IQ became a byword in everyday conversation.

This sudden popularization and publicity proved to be a mixed blessing in the development of a measuring instrument which was still in its infancy. Despite the fact that the existing intelligence scales were still very crude, they were too often accepted as a finished product and an infallible guide. Analysis of results and evaluation of techniques were subordinated to the more alluring occupation of classifying people. Occasionally psychologists themselves were guilty of overhasty generalization. Data on the various problems of differential psychology were being amassed in a rush. Sweeping conclusions were drawn—and quoted.

THE MEASUREMENT OF SPECIAL APTITUDES

The period between the two world wars witnessed many technical advances in psychological testing. The premature popularization of intelligence tests following World War I led to an inevitable reaction of skepticism among many laymen, as exaggerated initial expectations remained unfulfilled. In the meantime, psychologists were taking stock of this new tool, and an intensive phase of "testing the tests" was ushered in.

One of the principal results of such a critical study of psychological tests was a shift in emphasis from the exclusive use of intelligence tests to the measurement of *special aptitudes*. This shift is especially apparent in the testing of older adolescents and adults, in whom separate abilities are more clearly differentiated.[6] Among the manifestations of this growing concern of psychologists with special aptitudes is the tendency to report *sub-test scores* on the various parts of intelligence tests. For example, the American Council Psychological Examination,[7] administered annually since 1924 to the entering freshmen in many colleges throughout the country, is now scored so as to yield separate scores in the linguistic and quantitative parts. Prior to 1939 only a single score on the total test was regularly reported for each student.

Further evidence of the increasing recognition of the special aptitudes which enter into "general intelligence" is furnished by the *descriptive labels* now attached to many tests of the type formerly known as "intelligence tests." During the past two decades the designation "intelligence test" has been commonly superseded either by terms which describe more precisely what the test actually covers or by terms which connote preliminary classification only. An example of the former is the term "scholastic aptitude," now used to describe many "intelligence tests" which were found to measure principally those aspects of intelligence demanded by school work. An example of the latter practice is furnished by the Army General Classification Test (commonly referred to as the AGCT), which was developed in World War II to serve the same general functions as the Army Alpha of World War I (cf. 6, Ch. 11). This test provided a rough, prelim-

[6] This increasing differentiation of abilities with age will be discussed more fully in Chapter 14

[7] The full title of this test is American Council on Education Psychological Examination for College Freshmen, but it is commonly designated by the shorter title.

inary means of classifying the recruits into five Army Grades according to their general ability to learn the various duties required in military life.[8] It was prepared in four equivalent forms, each requiring about one hour, including preliminary instructions, a fore-exercise, and the test proper given with a 40-minute time limit. This test consisted of verbal, numerical, and spatial items, arranged in order of difficulty, and was given to every inductee who could read English. A later revision (AGCT–3), requiring about two hours, yielded separate scores in (a) verbal ability, (b) spatial comprehension, (c) arithmetic computation, and (d) arithmetic reasoning. This form, therefore, also illustrates the differentiation of total scores into sub-test scores discussed above.

The clearest indication of the emphasis upon special aptitudes is to be found in the large number of *special aptitude tests* which have been developed in recent years. Such tests are now regularly employed in individual guidance as well as in personnel selection. Although a general intelligence test is given as a preliminary instrument of classification for most jobs, such a measure is nearly always supplemented with more intensive testing in relevant areas. Many of these tests are custom-made for the specific job and are tested locally through a direct follow-up of a typical group of new employees. The Army and Navy also made much more extensive use of special ability tests in World War II than in World War I, when such tests were virtually non-existent (cf. 6, Ch. 11). Special "batteries," or combinations of tests, were constructed and assembled for pilots, bombardiers, range finders, radio operators, and scores of other specialized occupations of modern warfare. Tests of mechanical aptitude, clerical aptitude, motor dexterity, speed of reaction, visual and auditory acuity under various conditions, perception of distance or depth, and code learning are among the many special areas covered in these batteries.

It may also be of some interest to note that in a "poll of experts" (26), conducted in 1944, a representative group of psychologists in the testing field were found to be overwhelmingly in favor of continued development in the direction of aptitude testing. Of the 79 psychologists who replied, 55 expressed the opinion that "most will be accomplished if psychologists concentrate on measuring separate

[8] A similar test was prepared by the Navy. Known as the Navy General Classification Test, it consisted of 100 items similar to those in the original Army Alpha but expressed in naval terms and related to naval situations.

intellectual factors." Only 5 believed that the further development of testing should concentrate primarily on the measurement of general intelligence; 7 put the emphasis equally on general and special aptitude tests, and the remaining 12 gave no answer or an answer which could not be clearly classified into any of these categories.[9] It should not be concluded, of course, that this group of psychologists were dissatisfied with the current tests of intelligence. On the contrary, to the question, "In your judgment, how well do intelligence tests meet the practical needs for classifying people as to general mental ability in the army, in schools, and in industry?" over three-fourths of the group checked the reply, "Rather well, much better than is done without tests." In the comments following this question, however, it was again apparent that the intelligence tests were regarded as instruments of preliminary or approximate classification, which could profitably be supplemented by the measurement of special aptitudes.

PERSONALITY TESTS

The extension of testing techniques from sensori-motor and "intellectual" functions to emotional and social characteristics is also a relatively recent development. An antecedent of current personality testing may be found in Kraepelin's first use of the free association test on pathological cases and on persons who had been experimentally subjected to various influences such as fatigue, hunger, and drugs. Kraepelin (27) reported that all these agents increased the number of superficial associations. In 1894, Sommer (40) suggested that mental disorders could be differentiated by means of the free association test. The use of this test for a variety of purposes has persisted to the present day.

The most familiar personality tests, however, are those employing standardized *questionnaire or rating scale methods.* These methods were originally developed by Galton, Pearson, and Cattell for other purposes. The first systematic application of such techniques to personality testing is to be found in the Woodworth Personal Data Sheet (cf. 45, Ch. 5), an inventory constructed during World War I to detect neuroticism among soldiers. Although the armistice was signed

[9] One also wonders whether there is any significance in the fact that those respondents who emphasized the development of general intelligence tests were considerably older than those who gave precedence to the testing of special aptitudes!

before the final form of this questionnaire could be widely applied, it was subsequently used in army hospitals and in civilian testing. Several revisions and adaptations of the Woodworth Questionnaire have appeared, including forms especially suited for children, and for college students. Tests of the same type have also been developed for other social and emotional characteristics, such as introversion-extroversion and ascendance-submission. The adaptation of this technique to the measurement of interests and attitudes represents a still more recent ramification of personality testing.

In certain areas of personality, *performance tests* have been devised, the best known probably being the Hartshorne and May tests (21, 22, 23) for measuring character traits in school children. These comprise an extensive series of tests covering such behavior as cheating, lying, stealing, cooperation, persistence, and inhibition. All these tests are administered in everyday-life situations, including classwork, assigned "homework," athletics, or party games. The children are not aware that they are being tested or that their behavior can be detected or identified with them. Applications of these techniques to adults have been made from time to time in tests of such behavior characteristics as honesty, suggestibility, and persistence.

A type of personality test which has been gaining prominence in recent years comprises those tests which are commonly classified under the general heading of "*projective techniques*." In all such tests, the subject is given a task which permits of an almost unlimited variety of solutions. The assumption underlying these tests is that the subject will "project" into his performance his characteristic thoughts, worries, fears, attitudes, and other emotional responses. Several different tasks have been used for this purpose, including drawing, the arrangement of small toy objects to form a scene, extemporaneous dramatic play, the ranking of photographs in order of preference, and the interpretation of inkblots or of pictures.

The most extensively publicized of these projective techniques is undoubtedly the Rorschach Inkblot Test. In this test, ten cards are presented, each containing an irregularly shaped but symmetrical "inkblot." Five of the inkblots are in black and gray, and the remaining five in several colors. The subject examines each blot and reports all the associations and meanings suggested to him by the blot. His responses are then interpreted through a detailed scoring procedure which takes into account such features as responding to the whole blot

or to separate details, color responses, associations involving movement, and originality of associations, as well as the specific objects perceived. Another type of projective technique which has come to the fore following World War II is the sentence completion test. Attention was especially focused upon this technique as a result of its use in the O.S.S. Assessment Program. Such projective techniques are largely in an experimental stage: their procedure and scoring are not yet fully standardized, nor is their diagnostic significance conclusively established. The entire field of personality testing is at present in a formative stage and its technical development is far behind that attained in aptitude testing.

CURRENT TRENDS IN DIFFERENTIAL PSYCHOLOGY

Any summary of current trends in differential psychology must begin with a recognition of the rapid and uninterrupted growth in the sheer *number of psychological tests*. A glance at any of the bibliographies of mental tests which have appeared during the past decade yields ample evidence of such expansion. In the 1939 bibliography of mental tests and rating scales prepared by Hildreth (24), for example, 4279 titles are included; six pages are needed to cover the bibliography of bibliographies on tests. *The 1940 Mental Measurements Yearbook* (8), covering primarily paper-and-pencil tests published in English during approximately seven years, contains over 1500 titles. The third edition of this *Yearbook*, published in 1949, lists more than 600 additional tests which appeared between 1940 and 1947 (8). Such growth in tests has, furthermore, been multi-dimensional, testing being applied to more and more different aspects of behavior and employing increasingly varied techniques.

A second trend, especially apparent during the past decade, is the development of *methodological refinements* in both the construction and the application of tests. Not only are more rigorous techniques being devised for the selection and evaluation of test content and the establishment of norms, but there is evidence of a growing concern with the experimental design in which tests are used. Even the protracted controversies which have appeared in the recent literature on such issues as the effects of schooling (cf. Ch. 8) have helped to sharpen criticism and to focus attention on rigid standards of experimental control. The increasing practice of constructing tests to suit

the special needs of particular situations, as well as the practice of checking the diagnostic value of existing tests in the specific setting in which they are to be used, likewise reflect methodological progress. Mention may also be made of the trend toward more detailed analysis of test performance with reference to sub-tests and even individual items, as well as the study of the possible diagnostic significance of patterns of responses within a test.

A closely related trend, but one of sufficient importance to merit separate consideration, is the increasing frequency of *longitudinal studies*. The past decade has witnessed the completion of several major investigations involving ten- to twenty-year follow-ups of the same groups of individuals. A number of "growth studies" employing repeated psychological as well as physical measurements are now in progress, some beginning with individuals at birth (cf. Ch. 9). Follow-up studies of intellectually gifted children (cf. Ch. 17), as well as studies on the prolonged effects of special training (cf. Ch. 8) illustrate other applications of such an approach.

Differential psychology is also gradually emerging from the initial stages of description to embark upon an active search for underlying explanatory *principles*. Hitherto unrelated facts from various branches of the field are slowly being coordinated. More investigations are being designed with a view to studying the conditions under which individual differences develop or become modified.

Finally, increasing interest is being shown in the fundamental *nature of psychological traits* (cf. Ch. 15). Theoretical discussion of the concept of "trait" has flourished. Analyses of the specific components of "intelligence" and "personality" have multiplied. The combination of this growing theoretical sophistication with the development of methodological refinements should forecast a productive future for the differential approach to behavior phenomena.

CHAPTER 2

Basic Concepts of Psychological Testing

PSYCHOLOGICAL TESTING AND STATISTICS are the principal tools of differential psychology. As noted in the preceding chapter, it was the realization of the need for such tools in the study of individual differences which led Galton to the development of the first simple tests and to the establishment of the statistical laboratory which still bears his name. Subsequent developments in differential psychology closely paralleled the growth of the mental testing movement. Thus it should be apparent that an adequate understanding of the findings of differential psychology presupposes familiarity with at least the basic concepts of psychological testing and statistical method.

It is obviously beyond the scope of the present book to cover either of these fields. The reader is referred to any of the standard texts of psychological testing (cf., e.g., 7, 10) and of elementary psychological statistics (cf., e.g., 6, 9) for this purpose. Throughout the chapters which follow, however, some consideration will be given to the essential *implications* of the statistical techniques employed in the specific problems under discussion. Similarly, the present chapter will review the fundamental characteristics of psychological tests, without any attempt to survey specific available tests. Special attention will be given to the common procedures of test construction, since a knowledge of such procedures is essential to the proper interpretation of test scores.

BEHAVIOR SAMPLE

Every psychological test is essentially an objective and standardized measure of a sample of the individual's behavior. Any one test obviously covers only a small sample of the type of behavior which it is

designed to explore. In this regard, the psychologist's procedure is similar to that of, for example, a chemist who tests a shipment of iron or milk by analyzing one or more samples of it; from the directly measured characteristics of such samples the chemist then makes approximate deductions regarding the properties of the entire shipment. Similarly, when the psychologist wishes to measure the individual's vocabulary, arithmetic ability, or hand coordination, he observes the individual's performance with a limited number of words, arithmetic problems, or hand movements, carefully chosen to be representative of the total behavior segment he wishes to test.

Obviously both the number and nature of the items chosen for any specific test will determine the adequacy of coverage of that test. Thus an arithmetic test consisting of only five items would scarcely be expected to constitute an adequate sample of the subject's "arithmetic behavior." Nor could a satisfactory measure of such behavior be obtained from a test composed exclusively of multiplication problems. The diagnostic or predictive significance of any test depends upon the degree to which it serves as an indicator of a wider range of similar behavior. It is only in this sense that the psychologist can test behavior over and above that which he is directly measuring. For example, it might prove possible to predict an individual's ability to learn French from his performance in a one-hour test in the learning of an artificial language. If such were the case, we might say that the individual's "capacity" to learn French had been tested before he had even begun to study French. It is only through such sampling of relevant behavior, however, that a psychological test can predict "capacities." Contrary to certain popular notions, mental tests have no special powers for penetrating beyond observable behavior into a dark realm of hidden potentialities and latent aptitudes.

STANDARDIZATION

If the results of a psychological test are to have any value in predicting or diagnosing behavior, it is essential that the testing procedure be thoroughly standardized. The standardization of a test consists in the establishment of *uniform conditions* for administering the test to all individuals, as well as a uniform method for evaluating responses. It will be noted that this is simply a special application of the general requirement of controlled conditions in all scientific observation. The

one variable in a test situation is the individual who is being tested. Only if all other conditions are kept rigidly constant can the differences be attributed to the individual himself, who is the sole variable.

The layman generally has little realization of the degree to which uniformity of conditions must be maintained in a testing situation. It is largely for this reason that psychologists are skeptical of scores obtained by untrained examiners. It is not enough that the prescribed time limits be observed to the second and that the exact wording of instructions furnished in the test manual be followed. Such factors as the rate at which the directions are read to the subjects, the vocal inflections, pauses, and facial expressions which accompany them, and the exact placement of demonstration materials also affect the subject's performance. Altering any of these conditions may materially increase or decrease the difficulty of a test item for the particular person being tested. Such disturbances as undue fatigue or discomfort of the subject and distractions from persons walking about or from noises should obviously be avoided. The examiner must also make some effort to obtain the proper "*rapport*" with the subjects. This means that before beginning the test proper, he must arouse the subjects' interest and cooperation, make certain that he has their attention, and in every way insure that each subject will perform to the best of his ability. One of the advantages of individual over group tests is that with the former it is possible to establish rapport more fully and to maintain the subject's interest throughout the test.

NORMS

The Concept of Norms. The process of standardizing any psychological test includes not only the establishment of uniform procedure of test administration, but also the objective determination of norms. Without such norms it is impossible to interpret or evaluate the subject's performance on the test. By checking the subject's response on each item against a scoring key, the examiner determines the *raw score*. This raw score may be the total number of correct items, or the time required to complete a specified task, or some other objective measure of response appropriate to the content of the particular test. Such a score, however, has little or no meaning in itself. Psychological tests have no arbitrary or predetermined standards of "passing" or "failing", the individual's performance can be evaluated only in ref-

erence to the performance of other comparable individuals who have taken the same test. If, for example, an individual correctly identifies 75% of the words in a vocabulary test, such performance may be excellent, or just fair, or quite inferior. The question cannot be answered without reference to the norms for the particular test.

As its name implies, a norm represents the "normal" or average performance on a specific test. A test designed for 8-year-olds, for example, must first be administered to a large, representative group of 8-year-olds, in order to determine what is the average 8-year-old performance. If on this test the average 8-year-old completes 6 out of 15 problems correctly, then a raw score of 6 becomes the 8-year norm on this test. Once established, such a norm can be used in the future in evaluating the performance of any 8-year-old child who takes this test. In actual practice, norms provide not only the average score but also the relative frequency of varying degrees of deviation above and below the average, thereby permitting a more precise evaluation of scores throughout the entire range.

A word may be added regarding the interpretation of scores on *personality tests*. In such tests, there are generally no "right" or "wrong" answers. Consequently in the scoring key the responses are simply classified with reference to certain categories—such as "ascendant" or "submissive," for example—with no implication that either category is right or wrong. The concept of norm, however, is defined in essentially the same terms for personality tests as it is for intelligence or special aptitude tests. A personality test norm is an objectively determined average; it is not an ideal or perfect score, nor is it predetermined. For example, the norm on an emotional adjustment questionnaire might be represented by a raw score of 12 neurotic symptoms On tests of such characteristics as introversion-extroversion and ascendance-submission, the norm generally falls at a point midway between the two extremes. As in the case of all types of psychological tests, such norms are found by administering the test to a large sampling of people representative of the population on whom the test will ultimately be used.

Common Types of Norms. Norms have been variously expressed. Among the best known is the *mental age norm*, first developed by Binet. Age scales, such as the Binet and its derivatives, group the separate tests or items into age levels in terms of the performance of the subjects in the standardization group. Thus a test passed by the

average 7-year-old [1] is allocated to the 7-year level; one passed by the average 11-year-old, to the 11-year level. When the final standardized scale is administered to an individual subject, his score is expressed in terms of the highest year level he is able to reach. In actual practice, an individual rarely passes all tests at and below a given year level, while failing all those above it. A certain amount of "scatter" of performance over adjacent year levels is the usual pattern. In such a case, the individual is given additional partial credits (expressed in months or fractional parts of a year) in the determination of his final mental age.

In order to furnish a score which will be comparable at different ages, the mental age scores are generally transmuted into a relative measure or quotient, the most familiar of which is the Intelligence Quotient (IQ). This is simply the ratio of mental age to chronological age [2] $\left(IQ = \frac{MA}{CA}\right)$. Thus if a 10-year-old child obtains an MA of 10, his IQ will be exactly 100. When the MA is lower than the CA, the IQ will be below 100. IQ's above 100 signify that the individual's MA is higher than his CA, or that his test performance equals that of children older than himself.

An IQ of 80 or one of 120 represents the same degree of retardation or of acceleration, respectively, at all ages. This is not true of the mental age unit. Thus a retardation of one year in MA at age 4 is a more serious degree of backwardness than a retardation of one year at age 12. This follows from the fact that intellectual development, as measured by such tests, is more rapid during early life and exhibits a gradual slowing down with increasing age. There is a more marked difference between the tested abilities of the average 3- and 4-year-olds than there is between 11- and 12-year-olds.[3] Actually, the child who is one year retarded at the age of 4 will in general be three years retarded when retested at age 12. Under these circumstances, the IQ,

[1] In the actual construction of an age scale, the per cent of subjects of each age who pass tests assigned to that year level varies with age. This per cent must be greater at the lower ages than at the upper ages in order to yield the increasing variability of MA required for a constant IQ. In the Stanford-Binet, for example, the per cent of at-age passes drops from around 77 at age 2 to slightly below 50 at the average adult level. Even lower per cents of passes are used for the items at the superior adult levels in order to provide adequate ceiling.

[2] It is customary to multiply the resulting quotient by 100 in order to avoid decimals.

[3] Because of the inequality of successive mental age units, they cannot properly be averaged. This is especially true of mental ages which are far apart.

being a relative measure, will remain constant. In the present example, the child's IQ at age 4 will be 3/4, or 75; at age 12, it will be 9/12, or 75. Retest studies have shown that mental age deviations from the norm do in fact increase with age and that the IQ consequently tends to remain approximately constant, provided that the individual is not subjected to drastic environmental changes or other unusual conditions.[4]

One of the minor irritations to which psychologists are repeatedly exposed is to hear intelligence tests generically described as "IQ tests." Apart from the fact that the IQ is a score and not a type of test, it should be clearly recognized that all intelligence test scores are by no means IQ's. In fact, outside of a clinical situation, the majority of intelligence test scores are likely to be expressed in some other form. *Group tests*, for example, generally yield scores which do not lend themselves to expression in terms of IQ.[5] Similarly, tests designed for *adults* do not as a rule employ the IQ.

A brief consideration of the method for obtaining the IQ of an adult will demonstrate why the IQ concept is of little value in adult testing. In the first place, on an age scale such as the Stanford-Binet, the average individual's performance does not improve significantly beyond age 15. This means that the average 21-year-old or 30-year-old will do no better on the Stanford-Binet than the average 15-year-old. In order to compute an adult IQ on such a test, therefore, a divisor of 15 is used in place of the individual's actual chronological age. Such a procedure will yield an IQ of 100 for any adult whose MA is 15, i.e., whose test performance equals that of the average adult. It will be recalled, however, that the latest revision of the Stanford-Binet has three Superior Adult levels of increasing difficulty. If an individual were to pass all tests on the scale, through Superior Adult Level III, his MA would be 22 years and 10 months. Such a score is obviously a mental age only in an extrapolated sense and cannot be interpreted in terms of the original concept of mental age. A mental age of 8 represents the ability of the average 8-year-old, but a mental age of 22 does not represent the ability of the average 22-year-old—the average

[4] The problem of the constancy of the IQ will be covered more fully in Chapters 8 and 9.

[5] Specifically, a major requirement for the use of the IQ is that the extent of individual differences in MA increase systematically with age. This requirement is rarely if ever met by group tests.

22-year-old actually has a mental age of 15! In view of the above considerations, it is not surprising to find that the use of the IQ has been largely superseded in adult testing by one of the other types of scores to be discussed below.

A second type of norm is the *percentile norm,* based upon the per cent of individuals in the standardization group who fall below a given score. For example, if 68% of the standardization subjects score below 15 in an arithmetic reasoning test, then a raw score of 15 on this test corresponds to a percentile of 68. Anyone who completes 15 problems correctly, but no more, would receive a percentile score of 68. If we know that an individual has received a percentile score of 68 *on any test,* we can conclude that his performance excels that of the lowest 68% of the standardization group for that test. The 50th percentile on any test is obviously the midpoint, or median score. A zero percentile would signify a score as low as, or lower than, any of those obtained in the standardization group; a 100th percentile, a score higher than any obtained in the standardization group. The former need not indicate the failure to complete any items; nor does the latter necessarily represent a perfect score. Percentile scores should not be confused with the familiar percentage scores. The former are expressed in terms of *people,* the latter in terms of the number of *items* which the individual completes correctly. An ordinary percentage score is regarded as a raw score in psychological testing and is meaningless without reference to the norms.

Percentile norms furnish a convenient means of determining roughly the individual's relative standing in one or more tests. They are probably the most common type of norm employed with group tests. Certain cautions should, however, be observed in the use of percentile scores. The chief point to bear in mind regarding such scores is that they are essentially ranks and are therefore subject to whatever limitations apply to ranks. Percentile scores are measures of relative position, not of amount. For this reason, we cannot assume that successive percentiles represent equal differences in performance. In fact, we know that successive percentile scores near the mean, i.e., around the 50th percentile, correspond to much smaller ability differences than do percentile scores at the extremes of the range.[6] For example, the

[6] This follows from the normal distribution of performance and will be discussed further in Chapter 3.

distance between the 90th and 91st percentiles covers a much greater gap in performance than does that between the 50th and 51st percentiles. It follows from this inequality of units in different parts of the percentile scale that percentile scores cannot properly be averaged. This is particularly true at the extremes of the distribution, where the discrepancies in units become conspicuous.

A more precise and universally applicable technique for reporting test norms is in terms of *standard scores*.[7] Such scores use as their unit the standard deviation (SD or σ) of the distribution of scores of the standardization group. The standard deviation is a measure of the extent of variability or individual differences within a group. It is found by subtracting each individual's score from the group average, squaring each of these individual deviations, and then obtaining the square root of the average of these squares $\left(SD = \sqrt{\frac{\Sigma d^2}{N}}\right)$. Once computed for the particular group, the SD furnishes a convenient unit for indicating how far above or below the group average each individual falls. Let us suppose that the average raw score of the standardization group in Test A is 60 and the SD is 5. An individual with a raw score of 70 on Test A would then be two standard deviations above the average of this group ($70 - 60 = 10$; $10 \div 5 = 2$). Such an individual's *standard score* is said to be $+2$. Similarly, a raw score of 55 on Test A would correspond to a standard score of -1, and a raw score of 58, to a standard score of -0.4. A raw score of 60 would correspond to a standard score of 0, which always indicates the mean of the group.

A frequent practice, designed to avoid the use of decimals and negative quantities, is to convert the standard scores into a more convenient scale by adding an arbitrary constant and applying a constant multiplier to each score. A good illustration of this procedure is furnished by the scoring of the Army General Classification Test. In effect, the scores on the AGCT were standard scores which had been multiplied by 20 and added to 100. To put it differently, the raw AGCT scores were converted into standard scores in a distribution with a mean of 100 and an SD of 20, rather than the usual mean of 0 and SD of 1, which is implied in the simple conversion into standard scores.

[7] Also referred to as "sigma scores" and "z-scores."

Each of the five Army Grades into which the inductees were classified corresponds to one standard deviation unit in the AGCT distribution, as shown below (cf. 2, 11).

Army Grade	Distance from the Mean in SD units	AGCT Standard Scores
I	+1.5 and above	130 and above
II	+0.5 to +1.5	110 to 129
III	−0.5 to +0.5	90 to 109
IV	−1.5 to −0.5	70 to 89
V	Below −1.5	Below 70

It will be noted that a score of 100 was arbitrarily selected in this converted scale to correspond to the average performance of the standardization group.[8] Every 20 points above or below 100 represents one standard deviation above or below the group mean, respectively. Thus an individual with an AGCT standard score of 146 could be immediately recognized as being exactly 2.3 SD's above the group mean ($146 - 100 = 46$; $46/20 = 2.3$). Such converted standard scores combine the advantages of convenience with wide applicability and precise interpretation. These reasons explain why current tests are making increasing use of standard score norms.[9]

The Specificity of Norms. A requirement which is all too often overlooked in the application of norms is the comparability of the standardization group to the subjects on whom the test is to be used. It is not sufficient to know that norms were found on a large sampling. The nature of this sampling must also be taken into account in determining the uses to which the test may be put and in interpreting the scores. Very few tests are standardized on the general population. The Stanford-Binet and the AGCT probably come closest to these conditions. The former was standardized on a sampling which covered the school population in America quite adequately, although the selection of subjects at the lower and upper age ranges of the scale was not so representative; the latter furnishes norms on an unusually

[8] Unfortunately, because of this superficial similarity to the IQ scale, many laymen mistook the AGCT scores for IQ's, thereby contributing to the popular misuse of the latter term.

[9] For the statistically trained reader, standard scores (z-scores) should be differentiated from normalized standard scores (T-scores). Although the actual values obtained by the two methods do not usually differ very much, the assumptions underlying their computation are quite different. T-scores represent equal-unit scores in a normal distribution, while standard scores have the same form of distribution and the same inequalities of difficulty steps as the original raw scores.

extensive representation of American men of military age. Even these two samples, however, are obviously restricted. Age level is certainly one restricting factor in both, and other selective factors are undoubtedly present in milder degrees. The large majority of tests are considerably more restricted, their norms having been obtained on much more narrowly defined populations than was the case in these two tests. Thus one test may be standardized on college freshmen, another on groups of applicants for three or four specific jobs, a third on children from the fourth to the eighth grade of elementary school.

For many testing purposes, these more specific norms are of greater value. The test which undertakes to sample a more restricted population can, as a rule, do a more thorough job of sampling. At the same time, the comparison of an individual with a more clearly defined group to which he belongs permits a more significant interpretation of test scores. For example, it is usually of more practical importance to know how a college freshman's score compares with the average of freshmen in his own college than to know where the individual stands in relation to college freshmen in general. To compare such a score with the norm for the general population would be of still less value. Similarly, the personnel worker wants to know how a given applicant's test performance compares with the norms of applicants for the specific type of job to be filled, and preferably with norms obtained for this purpose in his own company.

It also follows that the scores from different tests—which have been standardized on different samples—are not directly comparable. Even when the populations for which the tests were designed are superficially the same, differences in the specific samples employed may significantly alter the meaning of the norms. Such differences were vividly illustrated in an analysis carried out in the Harvard Growth Study (4). Out of the total group of school children who were retested annually over a twelve-year period in this study, complete test records were available for 320 subjects. These subjects, all of whom had taken nine common group intelligence scales as well as the Stanford-Binet, served as the basis for a comparative study of the interpretation of an IQ derived from the different tests.

This analysis revealed considerable variation in the meaning of IQ's, not only from different tests, but also from different levels of the same test. Both of these fluctuations in IQ are attributable in large part to differences in sampling in the standardization groups. The

group used in the standardization of one test (or one age level of a given test) may have been somewhat brighter—or duller—than that used for another test or age level. Despite the efforts made to obtain representative samplings in the standardization of these tests, the groups actually used, in at least some of the tests, evidently fell short of this requirement.

For precise testing purposes, in which differences of even a few IQ points may be important, it is essential to allow for such variations in norms. This is particularly true in longitudinal studies, in which presumably equivalent tests (or levels of the same test) are administered at different times to the same subjects. In the Harvard Growth Study, the equivalence of IQ's on different tests was determined by reference to their relative position or percentile within the group under investigation. For example, the median IQ of the 320 children was 94 on one test, 102 on another, and 110 on a third. These three IQ's were then considered to be equivalent on the corresponding tests. Similarly, the 80th percentile of the group was represented by IQ's of 108, 118, and 124, respectively, on the same three tests. In this case, an IQ of 108 on the first test would "correspond" to one of 118 on the second and to one of 124 on the third test.

The variations actually found in this study are remarkably large. The authors concluded that, "an IQ of 100, which is commonly interpreted as indicating average ability and a position near the center of an unselected group, represents, on tests given for the first time, positions varying from the 19th to the 65th percentile . . . from one in the lower quarter of the group, representing an ability which is supposed to approximate dullness, to one near the upper third of the distribution, indicating brightness of a promising nature" (4, p. 134). It is thus apparent that an IQ—or any other type of test score—should be accompanied by an indication of the test upon which it was obtained. Such a score cannot be properly interpreted without full knowledge of the nature of the group from which the test norms were derived.

TEST RELIABILITY

The Meaning of Reliability. The concept of reliability is of fundamental importance in differential psychology. In all the varied applications of this concept, its common meaning is *consistency*. As applied

to psychological tests, reliability denotes the consistency of the subjects' scores upon retesting. If, for example, a given individual's IQ on a certain intelligence test is 135 on one day and drops to 86 upon retesting a few days later, the test obviously has very low reliability. Any one IQ found with this test would have little or no diagnostic value. Such retest fluctuations are known as the *error of measurement* of the test. Every test score will show some error of measurement, individuals rarely performing in identical manner on two occasions. Such marked changes as those in the above hypothetical example, however, would render a test useless for practical purposes. Changes of such magnitude might be the result of inadequate standardization of procedure, poor rapport, and other administrative conditions. Or they might indicate that the test is unduly susceptible to extraneous influences, such as weather conditions, or to minor emotional fluctuations of the subject, which would raise or lower the score on a particular occasion.

The Reliability Coefficient. The usual way of reporting the reliability of a test is by means of the reliability coefficient.[10] This is the coefficient of correlation between test and retest scores of the same group of subjects. The *correlation coefficient* (r) is a single numerical index of the degree of relationship or correspondence between any two sets of measures. This coefficient can vary numerically from +1.00, a perfect positive correlation, through 0, to −1.00, a perfect negative or inverse correlation. A + 1.00 correlation means that the individual receiving the highest score in one set of measures also receives the highest score in the other set, the one who is second best in the first is second best in the second, etc., each person's relative standing in the two measures being identical. A −1.00 correlation, on the other hand, indicates that the highest score in one measure is paired off with the lowest in the other, a corresponding perfect reversal occurring throughout the group. A zero correlation signifies no relationship at all between the two sets of scores, or the sort of arrangement which would result if the two sets of scores were shuffled and paired off at random. Perfect positive or negative correlations are very rare in actual practice, most of the coefficients falling on intermediate values.

The measure of correlation most commonly used in psychology is

[10] For a discussion of some of the implications of different techniques for the measurement of test reliability, cf. Thorndike (13).

the Pearson Product-Moment Correlation Coefficient. If, for example, we wish to correlate the scores on two administrations of the same test, A and B, by this method, we would find each person's deviation (or difference) from the average score on Test A and multiply it by the same person's deviation from the average score on Test B. The average of these products for the entire group is the correlation coefficient.[11] It is obvious that if those individuals who are above the average in Test A are all above the average in Test B, and those below in A are also below in B, then the products of the deviations will all be positive and the correlation will be positive. On the other hand, if most of the individuals who are above the average in A (i.e., positive deviations) are below the average in B (i.e., negative deviations), then the products of these deviations will be negative and their average will be negative, thus yielding a negative correlation.

The computation of reliability coefficients is one of the many uses to which the correlation coefficient is put in psychological testing. Most tests in current use have reliability coefficients in the .80's or 90's. Reliabilities which fall appreciably short of these values do not meet the standards of consistency needed for most testing purposes.

The Role of Behavior Fluctuations. If we regard test reliability as an index of the *consistency of the test as a measuring instrument*, then perfect reliability is not at all inconsistent with fluctuations of responses on different occasions (cf. 1). A discrepancy in score on successive retests may simply mean that the test is serving its function as an accurate and sensitive index of actual changes in the subject. To take an example from a different field, one does not measure the reliability of a thermometer by comparing temperature readings on different days. The thermometer may be perfectly reliable and still give very different readings on successive days. Such fluctuations in daily temperature readings would be of interest if one wished to determine the reliability of an estimate of daily temperature in a given locality, from a single day's reading. It would, in other words, indicate the consistency of the temperature, not of the thermometer. Similarly,

[11] Since it is essential that the two deviation to be multiplied be in the same units, the actual computation of r involves the transmutation of all scores into standard scores. Several formulas are available for facilitating the computation of this coefficient, all of them automatically performing this transmutation. The most familiar formula is: $r = \frac{\Sigma xy}{N\sigma_x\sigma_y}$, in which Σxy is the sum of the products of the deviations discussed above, σ_x and σ_y are the standard deviations of the two sets of scores, and N is the total number of cases in the group. Several other formulas have been developed which permit further computational short-cuts.

the discrepancies in score on different occasions may simply show how variable the subjects are in the functions tested.

This problem is not especially serious in the measurement of aptitudes, since the individual's abilities in arithmetic, vocabulary, mechanical comprehension, motor coordination, and the like are not likely to alter appreciably from day to day. Hence fluctuations in such tests can be attributed to the "unreliability" of the measuring instrument and steps can be taken to improve or replace the test. If the interval between successive retests is one of several years—or even several months in the case of young children—then, of course, the changes in score cannot be attributed solely to the tests. Similarly, if the subject has undergone drastic changes in environment or in physical condition during the interval between testings, sharp rises or drops in test score may occur which have no bearing upon the reliability of the test.[12] In the absence of any unusual circumstances and with intervals of only a few weeks or days between retests, however, it is probably safe to expect aptitude test scores to remain approximately constant. To demand high retest correlations in such cases is therefore justified.

In the field of personality testing, on the other hand, it is reasonable to suppose that an individual's attitudes, dominance, self-confidence, and the like may vary appreciably even over short intervals. The susceptibility of such characteristics to the individual's experiences immediately preceding the test cannot be ignored. This is not meant to imply that personality characteristics do show daily fluctuations, but only that we cannot legitimately assume their constancy, and thus cannot ascribe all sources of variation in score, *ipso facto*, to imperfections of the test. It is apparent that the retest method of determining test reliability does not lend itself very well to personality tests, since a low retest correlation in such a test would be ambiguous.

The Sampling of Content. When a test is of such a nature that the subjects can recall some of their responses from the first to the second testing, or when an appreciable practice effect occurs in the course of the testing, then two *equivalent forms* of the test are administered rather than repeating the identical test. Such equivalent forms are composed of different items chosen as samples of the same abilities. The correlation between such equivalent or parallel forms thus depends both upon the day-by-day consistency of the scores and upon

[12] For a discussion of relevant material, see Chapters 8 and 9.

the degree to which each form actually samples the entire behavior area which is being measured. For example, if a list of 20 carefully selected words is used to sample the individual's vocabulary knowledge, how far will the subject's score change when he is given another vocabulary list chosen in the same manner but containing 20 different words? Obviously the longer the list of items the higher, in general, will be the consistency of the subject's performance on the second, equivalent list. Other things being equal, a longer test can sample the behavior in question more thoroughly than a shorter test. It is, in fact, possible to predict by means of a statistical formula [13] approximately how much the reliability coefficient of a test will rise when the test is lengthened by specified amounts.

Test reliability is sometimes defined exclusively in terms of the adequacy with which the test samples the behavior under consideration. Such a definition is implied by the common practice of computing reliability by the *split-half technique*. This type of reliability coefficient, sometimes known as the coefficient of internal consistency, is universally applicable to all types of tests and is undoubtedly the most widely used. Not only does it avoid the recall of items on retests, but it also rules out any general effects of practice, fatigue, or similar cumulative factors. The split-half reliability is likewise uninfluenced by the possibility of "true" daily fluctuations, as in the case of personality tests discussed in the previous section. The procedure consists essentially of correlating two sets of scores obtained during a single administration of a single form of the test. Perhaps the most common form of the split-half procedure is the "odd-even technique," in which each subject's score on the odd items is correlated with his score on the even items. In this way, neither score has any appreciable advantage or disadvantage in terms of adaptation, practice, fatigue, boredom, difficulty level of items, or any other condition which may vary progressively during the test period. Other ways of dividing the test can, however, be employed. In a speed test, for example, the performance during the first and last quarters of the test period is usually combined to yield one of the two half-scores, while the performance during the two middle quarters determines the other.

It will be noted that the "correlation of halves" thus obtained actually shows the reliability of only half the test. If, for example, the entire test consists of 100 items and the reliability coefficient is com-

[13] Known as the Spearman-Brown formula—cf., e.g., Garrett (6), pp. 387–390.

puted by correlating scores on two sets of 50 items each, the obtained reliability will obviously be lower than that of the entire test. For this reason, it is customary to estimate from the correlation of halves what would be the reliability coefficient of the whole test, i.e., of a test double in length. The formula discussed above for determining the effect of lengthening a test upon its reliability coefficient is used for this purpose; the estimated reliability is accordingly designated the "Spearman-Brown reliability coefficient."

It is apparent that the three principal methods for determining the reliability of a test differ not only in the range of situations to which they are applicable, but also more fundamentally in the aspect of reliability which they measure. The retest method indicates the degree of day-by-day consistency of performance; the equivalent-form method combines day-by-day consistency with adequacy of behavior sampling; the split-half method is based only upon adequacy of behavior sampling.[14]

The Effect of Range. The reliability coefficient, in common with all correlation coefficients, is influenced by the range of scores within the group on which it is computed. In general, the reliability coefficient of a test will be lower when found on a group of a single age level than when computed on a group of varying age. The latter, more heterogeneous age group will exhibit a wider range of scores on most tests than will a group homogeneous in age, and will consequently yield higher correlation coefficients. The age range of a group is one of the most frequent sources of discrepancy in correlation coefficients from sample to sample, and its influence should always be taken into account in evaluating the size of an obtained correlation. Any other factors which serve to increase or decrease the heterogeneity of a group may also raise or lower the size of a correlation coefficient.

It thus follows that any given test has not one but many reliability coefficients. The same test will yield a high reliability coefficient in a relatively heterogeneous group, a much lower one in a more homogeneous group. A measure which is relatively independent of the range of scores is the *standard error of measurement*[15] (σ_{meas}). In

[14] A fourth method which has come into prominence in recent years is the "method of rational equivalence" developed by Kuder and Richardson (8). It is also basically a measure of internal consistency and therefore an index of the adequacy of behavior sampling. Its computation takes into account the intercorrelation of individual test items.

[15] Or the probable error of measurement, which can be obtained by multiplying the standard error of measurement by .6745 ($PE_{meas.} = .6745\sigma_{meas}$).

the computation of this measure, both the reliability coefficient and the standard deviation of the scores of a particular group are employed. The result will be in the same units as the test scores, and indicates the amount of "error" introduced into the score by the use of an imperfect measuring instrument. In interpreting the σ_{meas}, we may say that the chances are approximately 2:1 that the obtained score does not differ by more than the amount of the standard error from the individual's "true" score, i.e., the score he would have obtained on a measuring instrument with perfect reliability. To take a specific illustration, if a child's IQ on a particular test is 113 and the σ_{meas} of this test is 5 points, then the chances are 2:1 that the child's "true" IQ is between 108 and 118 ($113 - 5 = 108$; $113 + 5 = 118$).

VALIDITY

The Concept of Validity. The validity of a psychological test is the degree to which the test succeeds in measuring, diagnosing, or predicting that area of behavior which it sets out to measure. In order to determine such validity, it is necessary to have an independent *criterion measure* of the behavior under consideration. For example, in validating a series of tests of musical aptitude, subsequent performance in music schools was the criterion employed. A test designed for the selection of taxi drivers would be validated against actual job performance of a typical group of applicants. A scholastic aptitude test for college freshmen would be checked against the students' grades in college courses. In all these situations, a representative group, selected for validation purposes, is given the test and then followed up to determine each individual's actual performance in the areas being tested.

The follow-up thus yields a direct measure of that which the test is trying to predict through a small sample of performance. In this lies the answer to the apparent paradox of test validity. It might have been argued that if we need an independent and thoroughly reliable measure of that which the test seeks to measure, then why do we need the test at all? The purpose of the test, however—once it has been validated on the trial group—is to *predict* within a short testing period that which would otherwise have been discovered only through a long and wasteful period of direct observation. If all applicants for

a job were indiscriminately hired, the sheep would eventually be separated from the goats through actual success or failure on the job. Within a period of several years, most of the incompetent workers would probably have been eliminated. Such a process of selection would obviously be absurdly costly both to employer and employee. Similarly, all students who wished to enter medical school might be admitted, on the undoubtedly correct assumption that those lacking the proper qualifications would eventually "flunk out." It is the primary object of psychological tests to approximate, in advance, the type of judgment which would otherwise require a virtually prohibitive trial period. Once the validity of a test has been established, there is, of course, no longer any need for a criterion measure on the subjects with whom the test will be used.

The validity of a test is most commonly reported in terms of the *validity coefficient*, i.e., the correlation between test scores and criterion measures in the validation group. All the factors which may affect the size of a correlation coefficient should obviously be taken into account in evaluating such a validity coefficient. Thus the size and nature of the group upon whom the correlation was computed must be ascertained. A validity coefficient obtained on a small number of subjects is of little significance, since the correlation may vary widely when a different sampling is employed. The type of subjects upon whom validity was determined should be similar to those on whom the test is used. It is now a growing practice to redetermine the validity of published tests on the population with which the test is to be used. Not only norms, but also the validity of a test, are specific to the population. A test may be a good measure of intelligence for machinists and a poor one for college freshmen. Or the same test may actually prove to be a satisfactory test of a different function when applied to a group unlike the original validation group.[16]

Also relevant is the consideration of the range of ability represented by the group upon whom the validity coefficient is computed. This is the same problem which was discussed in connection with the reliability coefficient and it can be met by a similar solution. In place of the standard error of measurement, we can now compute the *standard error of estimate*[17] (σ_{est}). This is a measure of the amount of

[16] For example, a well-known test of clerical aptitude proved to be a poor instrument for the selection of office clerks in a particular company, but was successful as a means of selecting a certain type of routine factory worker.

[17] As usual, this can also be expressed as a *probable error*: $PE_{est} = .6745\ \sigma_{est}$

error which is introduced by virtue of the fact that the individual's performance is predicted or estimated from the test score rather than being directly measured through the criterion. The σ_{est} is based upon the SD of the criterion measure and the validity coefficient of the test. It is expressed in the units in which the criterion is measured. For example, if scholastic grades are to be predicted and if such grades are expressed on a percentage basis, then the σ_{est} will be in percentage units. Thus if a student's predicted grade on such a scale is 78 and the σ_{est} of the test from which the prediction was made is 4, then the chances are 2:1 that the student's actual grade will lie between 74 and 82 ($78 - 4 = 74$; $78 + 4 = 82$).

In a number of situations, tests are used not to predict the exact performance or relative position of each individual, but rather to select those individuals who are most likely to reach or exceed a certain minimum standard of achievement. This is particularly true of the industrial use of tests, in which the test serves as a "*screening*" device, and "*cut-off scores*" are often of more interest than validity coefficients. The validity of a test can be most clearly determined in such situations by the method of *contrasted groups*. For example, a group of applicants who have taken the test under consideration are hired and followed up for a year or longer. At the end of this trial period, each individual is classified as a "success" or "failure" in terms of whatever practical standard the company employs for evaluating job performance. The initial test scores of these two contrasted groups are then compared: the greater the difference in such scores between the two groups, the greater the validity of the test for the purpose of selecting successful applicants. From an examination of the distribution of scores made by the two groups, a "cut-off" point or minimum score is set at a point which would eliminate the largest possible proportion of "failures" and at the same time exclude the smallest possible percentage of "successes." [18]

Illustrations of test validation taken from the military use of tests in World War II are shown in Figures 1, 2, and 3, the first two dealing with the validity of the AGCT and the third with the validity of a special battery of tests designed for the selection of pilots. Figure 1

[18] Knowing the validity coefficient of a test, the per cent of applicants who will need to be hired, and the per cent of "successes" on the particular job prior to the use of the test, it is possible to compute the per cent of those selected through the test who will succeed (cf. 12). In this way, the net gain in selection accuracy attributable to the use of the test can be determined.

indicates the extent to which the AGCT predicted the success of officer candidates, the criterion being the actual commissioning of the men at the completion of their officer training course. The data

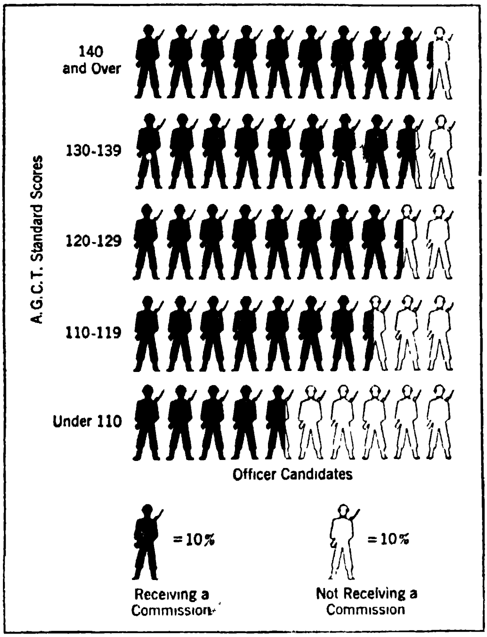


Fig. 1. Predicting Success of Officer Candidates from their Scores on the Army General Classification Test. (From Boring, 3, p. 242.)

of 5520 officer candidates from 14 schools reveal a fairly close degree of correlation between criterion and test score. Thus of the men scoring 140 or higher on the AGCT, over 90% received their commission. At the other extreme, less than 50% of those scoring below 110 succeeded in obtaining the commission, although they had gone through the same training course. Figure 1 shows very clearly the reason for setting 110 as the cut-off point for subsequent admissions to officer training courses. While over half of those scoring below 110 failed, the number of failures dropped sharply to less than a quarter in the group scoring 110–119.

Similar data for tank mechanics are given in Figure 2. Grades obtained in a tank mechanics course constituted the criterion for this group. Subdividing the group into the five Army Grades on the

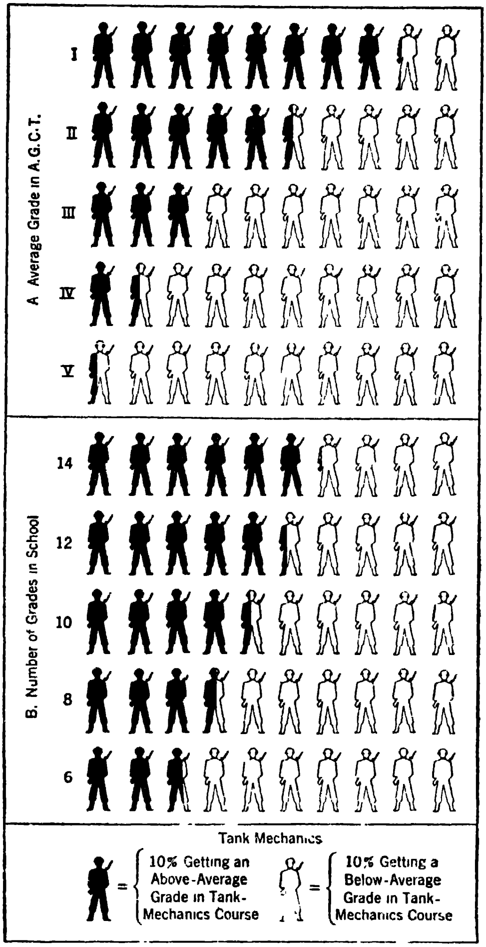


Fig. 2. Predicting Success in a Tank Mechanics Course from AGCT Score and from Schooling. (From Boring, 3, p. 251.)

AGCT, we find an even closer correspondence between test score and criterion than was found in the officer group, a fact which is undoubtedly attributable in part to the greater heterogeneity of the tank mechanics. It will be noted that among the men scoring below 70 on the AGCT, less than 10% received an above-average grade in the tank mechanics course. This percentage rises in each successive Army Grade until we find over 80% of the men in Grade I receiving

an above-average grade in the course. For comparison, the data on schooling have been checked against the same criterion of success in the tank mechanics course, with the results shown in section B of Figure 2. Some correspondence exists here also: among the men who had completed only six grades in school, less than 30% did better-

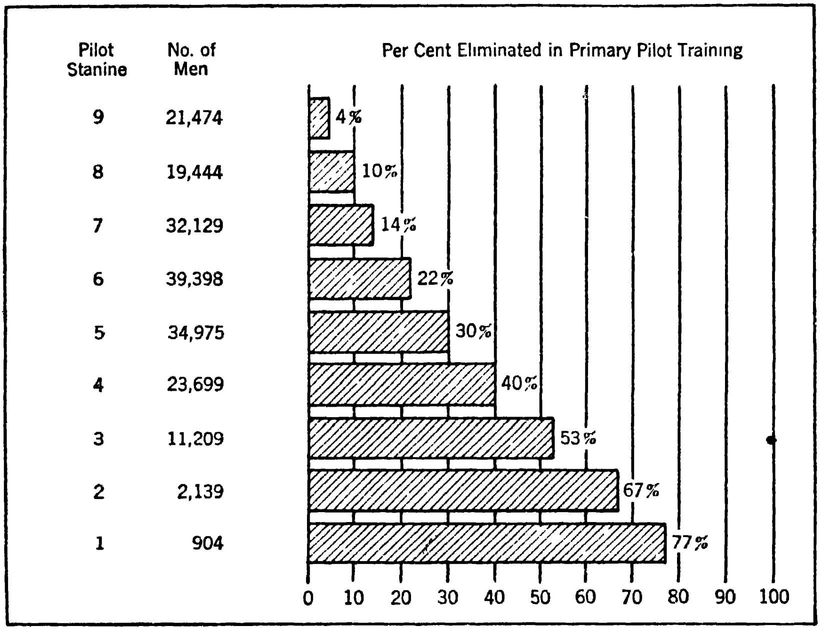


Fig. 3. Validity of a Pilot Selection Battery. (From Flanagan, 5, p. 58.)

than-average work in the tank mechanics course, this percentage rising slowly but consistently until it reaches slightly more than 60% in the group that had had two years of college training. It is apparent, however, that success in the tank mechanics course correlated more closely with AGCT score than with schooling.

The data reproduced in Figure 3 show the effectiveness of a series of tests used by the Army Air Forces in the selection of pilots. The object of the tests was to predict pilot success prior to any flying training. The major criterion employed for validation was completion of primary pilot training, although some corroborative data on subsequent performance in advanced training was also obtained. Through an initial analysis of the requirements of the pilot's job, together with

preliminary follow-up studies, a battery of tests was assembled, which consisted of 6 apparatus tests of coordination and speed of decision, and 14 paper-and-pencil tests of intellectual abilities, perception and visualization, and certain personality characteristics. The scores on each of these tests were weighted so as to give the best possible estimate of the criterion. The sum of the weighted scores was expressed in terms of a nine-point scale. The term "stanine" (standard nines) was coined for the units on this scale, a stanine of 1 representing the lowest level of estimated pilot aptitude and a stanine of 9 the highest.[19] Reference to Figure 3 shows the degree to which this battery succeeded in separating the successes from the failures in primary pilot training, in a group of 185,367 men. The horizontal bars indicate the proportion of men in each stanine who were eliminated because of flying deficiency, fear, or at their own request. It will be seen that only 4% of those in the top stanine were eliminated, in contrast to 77% in the lowest stanine. The per cent of failures follows a uniform progression between these two extreme groups. As a result of such investigations, the cut-off score for admission to pilot training[20] was subsequently set at pilot stanine 7.

The Validation of General Intelligence Tests. The validation of special aptitude tests is a relatively clear-cut procedure since the criterion is usually definable in fairly specific terms. Such is also the case when validating general intelligence tests for use as instruments of preliminary classification or as part of a battery for the prediction of performance in a particular occupation, course of study, and the like. Most intelligence tests, however, are still constructed for the relatively general and vaguely defined purpose of measuring "intelligence," without reference to any specific situation. The problem of finding a suitable criterion for such tests presents more serious difficulties.

A number of special criteria have been developed for the validation of tests of general intelligence, more than one criterion frequently

[19] "Stanines" are normalized standard scores, or "T-scores" (cf. footnote 9). The lowest 4% of the distribution received a stanine of 1; the next 7%, 2, the following 12%, 3; the next 17%, 4, and the following 20%, 5, which corresponded to the mean of the distribution. Corresponding percentages above the center of the distribution (17, 12, 8, 4) were assigned stanines of 6 to 9, respectively.

[20] Similar batteries were constructed to yield stanines for navigators and other AAF groups. The specificity of aptitudes was vividly demonstrated by the observation that men with high navigator stanines, for example, often had pilot stanines as low as 3, 4, or 5.

being applied in the evaluation of a single test. For age scales such as the Stanford-Binet, as well as for preschool and infant tests, *age differentiation* is a major criterion. The degree to which the test discriminates between successive age groups, as well as the correlation between chronological age and score, are taken as indices of validity. It should be noted that the satisfaction of such a criterion indicates merely that the test measures behavior characteristics which tend to increase with age under existing conditions and in the type of environment in which the test is standardized.

A second, commonly used criterion consists of teachers' ratings, school grades, or other indications of quality of *academic achievement*. It is because so many current intelligence scales have been validated chiefly against school achievement that they are frequently described as tests of scholastic aptitude. *Ratings by supervisors*, such as employers, shop foremen, army officers, and the like often serve as criteria for adult tests. *Amount of education* is sometimes introduced as a criterion on the assumption that the successive rungs of the educational ladder serve as selective factors, progressively eliminating those less able to profit from the more advanced types of instruction. With an extensive system of public education, such as is available in America, this assumption is partly correct, although at the higher educational levels other factors besides ability undoubtedly affect survival.

Various applications of the method of *contrasted groups* are also to be found. The comparison of the scores of persons in different occupational levels is an example of such a method. Another illustration is the comparison of the scores of unselected school children with those of institutionalized feebleminded subjects of the same age. In these instances, the criterion is ultimately based upon the composite demands of everyday life situations which determine survival in various occupations or in a normal, non-institutional environment. A closely related validating technique is to correlate[21] test scores with psychiatrists' diagnoses as to whether or not the individual should be institutionalized for mental deficiency. Unless such diagnoses are based upon a prolonged observation period, however, this criterion

[21] By means of the biserial coefficient of correlation, since the diagnoses are in a twofold category and the test scores in a continuous distribution. Such correlations have been reported, e.g., in the validation of the Wechsler-Bellevue Intelligence Scale.

may itself be no more valid than the test and thus serve no purpose in the process of validation.

Frequently, *correlations with other intelligence tests* are reported as validity coefficients. The Stanford-Binet, for example, has often served as such a criterion. This procedure is justified only when the new test is a short and relatively crude instrument introduced as a practical time-saving device. It obviously assumes that the new test can do no better than approximate the results of the earlier test. For some tests, however, such correlations are reported not so much as validity coefficients but simply as a rough indication that the new test is measuring approximately the same general area of behavior as other current tests designated by the same name. In such cases the logic of the situation demands that the inter-test correlation be moderately high, but not too high. An unduly high correlation between such tests would indicate needless duplication of effort, since if the two tests are so nearly the same, there is no point in introducing the second one.

Finally, we may consider the method of *internal consistency* whereby the total score in the test itself is used as a criterion. Individual items for a test are often chosen on the basis of their agreement with the total score. For example, items on the Stanford-Binet were selected in part by comparing performance on each item with IQ on the entire scale. If, for instance, a given item was passed by approximately the same per cent of subjects in the lower and upper IQ levels, it obviously was failing to discriminate between individuals who differed in those characteristics measured by the test as a whole. On this basis, such an item would be eliminated from the scale. The degree to which the items finally selected correlate with the total score may then be cited as evidence of validity. Another application of this method is the selection and validation of the separate tests in a battery, in terms of the correlation of each test with the composite score on the entire battery.

The method of internal consistency falls in the borderland between validity and reliability. In so far as it does not depend upon an outside criterion, it does not, strictly speaking, yield a measure of validity. It may be argued, however, that any index of the adequacy with which a given behavior area is sampled is relevant to the concept of validity. It is certainly true that a test cannot be very valid if its

items do not adequately sample any one behavior characteristic, i.e., if the subject's performance is inconsistent from item to item and his score varies widely with the addition or deletion of a few items. On the other hand, although a test may measure a particular behavior area with a high degree of consistency, and may have sampled it very fully, the behavior tested may not be that which the test purports to measure. It would be quite possible, for example, to devise a test which showed a high degree of internal consistency, but which did not differentiate between normal and feebleminded subjects. If such a test had been labeled an "intelligence test," it obviously would not be valid, despite its high internal consistency. This technique is thus of value only when used in conjunction with further validation by outside criteria.

The Validation of Personality Tests. The validation of personality tests presents even more of a problem than does the validation of intelligence tests because of the difficulty of finding a satisfactory independent criterion of most personality characteristics. A number of the techniques employed with intelligence tests have, however, been adapted for use in the validation of personality tests, with a moderate degree of success. One such technique is the correlation of test scores with *ratings* by associates, teachers, job supervisors, and others who may have had an opportunity to observe the subject over an adequate period of time. In general, such criterion ratings for personality characteristics should be made by more than one observer, in order to guard against individual bias and other idiosyncrasies of the raters. Similarly, care should be taken to insure that the raters have had "trait acquaintance," i.e., that they have had the opportunity to observe the subjects in those specific aspects of behavior which are covered by the test. Correlation with *psychiatric diagnosis* has been employed in validating certain tests of emotional maladjustment. As in the case of intelligence tests discussed above, such a procedure is satisfactory only when the criterion itself is based upon a careful and prolonged follow-up, rather than upon a cursory psychiatric examination which may be no better than the test being validated.

Correlations with *other personality tests* have sometimes been reported as an index of validity. This again presupposes that the criterion itself has previously been established as valid. It is best adapted to the validation of tests which are introduced as abridged, time-

saving versions of longer tests. For example, the Bernreuter Personality Inventory was originally designed to yield four separate scores within a 15–20-minute testing period, each of which was an approximation of the score on a different, previously constructed test. Through this test a rough estimate was obtained of the subject's score in neuroticism, introversion, dominance, and self-sufficiency. Obviously in this case the correlations of the Bernreuter scores with the scores on the four separate tests from which it was derived would be relevant.

The method of *contrasted groups* is one of the most common ways of checking the validity of a personality test. For example, occupational group may be the criterion, as when a test of extroversion or sociability is given to, let us say, salesmen and mechanics. If the scores of the former group are clearly higher than those of the latter, some evidence will thereby have been furnished for the validity of the test. Delinquent and non-delinquent children have occasionally been used in a similar way to test the validity of certain character tests. Or the scores made by neurotics under treatment can be compared with those of a matched group of normal persons who have never been under psychiatric care. In connection with tests of neuroticism or emotional instability, the *frequency of a response* in a normal group is a further check. Thus if a particular behavior characteristic occurs in a large per cent of normal persons, it cannot by definition be an "abnormal" response.

A relatively large number of personality tests, especially those of the questionnaire type, have relied exclusively or primarily upon the method of *internal consistency*. For example, the 25% most introverted and the 25% most extroverted subjects in the validation group are first selected on the basis of their total scores on a preliminary form of an introversion-extroversion test. The responses of the two groups on each item are then compared. If a supposedly "introverted" behavior item occurs more often among the extroverted group than among the introverted, then such an item is discarded as not being properly diagnostic. If it occurs with about equal frequency in both groups, it is neutral or irrelevant and should likewise be discarded. To be retained, an "introvert" item must occur with a significantly higher frequency in the introverted than in the extroverted group. This method is subject to the same limitations discussed in

connection with its use in the validation of intelligence tests. The fact that it has often been the only method for checking the validity of personality questionnaires has led to considerable skepticism regarding the behavior characteristics which these tests were actually measuring.

THE QUESTION OF "CAPACITIES"

In closing this brief survey of some of the major problems of psychological testing, a word should be added regarding the relationship of tests to the commonly misused concept of "capacity." The original aim of the mental testers was the measurement of the individual's "capacities," or "potentialities," of behavior development, as distinguished from his present skills and information. The measurement of the latter would have been a relatively simple task. If we want to ascertain whether an individual is proficient in many languages, for example, we need only to examine his knowledge of all languages with which he claims familiarity. But if we want to know whether this individual can learn languages easily, whether it would be worth the effort to teach him, or whether he should consider a vocation which demands a mastery of several languages, then we are faced with a much more difficult problem. This is the type of problem with which mental testers have tried to cope.

If one is to determine what the individual *can* do rather than what he has already accomplished, it has been argued, it is necessary to "rule out" in some way the differences in formal or specialized training among different individuals. This is usually attempted either by presenting material which is equally unfamiliar to all or by the reverse procedure of utilizing only material common to everyone's experience. Frequently the two methods may be combined in different items of a test, or even in the same item, as in the use of familiar material in a novel and unusual manner.

Such a procedure is a practicable one and will yield usable information, provided that due cognizance is taken of its assumptions and limitations. In the use of either "familiar" or "unfamiliar" material, it is necessary to ascertain whether the material is actually familiar (or unfamiliar), to an approximately equal degree, to all the subjects being tested. When given to persons from different national or cultural groups, or from widely differing economic, social, or educational back-

grounds, psychological tests do little more than reflect the varied backgrounds of the subjects.

No psychological test has any mysterious power in itself whereby it can strip the subject's behavior of the accumulated effects of his reactional biography, and reveal his original, carefully insulated "potentialities." In mental testing, the terms "potentiality" and "capacity" can be used meaningfully only in the sense of prediction of *subsequent behavior* from present behavior. The prediction may likewise cover a *wider range of behavior* than that included in the test, if such a prediction is proved to be valid. But the starting point of such predictions is always present behavior, not anything projected back into some hypothetical pristine state.

CHAPTER

3

Nature and Extent of Individual Differences

POPULAR OPINION FREQUENTLY CLASSIFIES PEOPLE in reference to the possession or non-possession of certain traits. Thus one individual is said to have a talent for music, another for painting, a third for mathematics, a fourth for organizing people. Such a characterization, however, results from purely practical considerations. In order to choose music as a vocation, or even as a serious avocation, for example, an individual must have a certain minimum of musical talent; if his degree of musical ability falls below that minimum, he is not regarded as "a musical person." Moreover, in our society we are accustomed to characterizing the individual in terms of his *outstanding* assets and liabilities, and simply ignoring the traits in which he rates close to the norm. Hence we label Mr. Jones a violinist, Miss Smith a skater, and Mr. Doe a thief. We do not ordinarily characterize Mr. Doe as a mediocre skater, Miss Smith as a relatively poor violinist, or Mr. Jones as an "average honest man!" [1] Qualitative distinctions of this sort are made in practice and are based on arbitrary or socially determined criteria or limits.

Actually, however, every individual can be described along a continuous scale in any behavior category. In other words, individuals do not fall into sharply divided types, individual differences are rather a matter of *degree*. It is in this sense that individual differences are said to be *quantitative* rather than qualitative.

To be sure, it might be argued that there are certain characteristics which a person may either have or not have, and that in this respect we may speak of qualitative differences. The classical examples are such sensory handicaps as loss of vision or hearing. Here, it would

[1] The very fact that we have a word for "thief" but no word for "average honest man" is a further illustration of the same point.

seem, are traits characterized by presence or absence: a person can see or he cannot see, he can hear or he cannot hear. This, too, turns out to be a purely conventional and practical distinction. Anyone who has visited a school for the blind knows that there are many degrees of blindness, and that not all those classified as blind are totally blind. The everyday working definition of blindness is any *degree* of visual deficiency too serious to permit normal activity. The same is obviously true of deafness and any other sensory disorder. Between the empirically established "normal" vision or hearing and what is classed as blindness or deafness there is to be found a continuous gradation of minor deficiencies. It should be added that the existence of a trait in zero degree, as in total blindness, is not inconsistent with the quantitative view of individual differences. The latter implies only that there be intermediate degrees rather than simple presence or absence.

THE DISTRIBUTION OF INDIVIDUAL DIFFERENCES

Since individual differences are quantitative in the above sense, we may now ask how the varying degrees of each trait are distributed among people. Are individuals scattered uniformly over the entire range or do they cluster at one or more points? What are the relative

TABLE 1 *Frequency Distribution of Scores of 1000 College Students on a Simple Learning Test*

(From Anastasi 3, p. 31)

Class-Interval	*Frequency*
52–55	1
48–51	1
44–47	20
40–43	73
36–39	156
32–35	328
28–31	244
24–27	136
20–23	28
16–19	8
12–15	3
8–11	2
	N = 1000

frequencies with which different degrees of a trait occur? These questions can best be answered by an examination of frequency distributions and frequency graphs.

Like all statistical devices, the frequency distribution is a means of summarizing and organizing quantitative facts in order to facilitate their treatment and reveal significant trends. Scores on a test, or any other set of measures, are grouped into class-intervals, and the number of cases falling within each interval is tabulated. An example of a frequency distribution is given in Table 1. This shows the scores of 1000 college students on a simple learning test. The scores range from 8 to 52 and have been grouped into class-intervals of four points. The advantages of such a table over a list of 1000 individual scores are obvious.

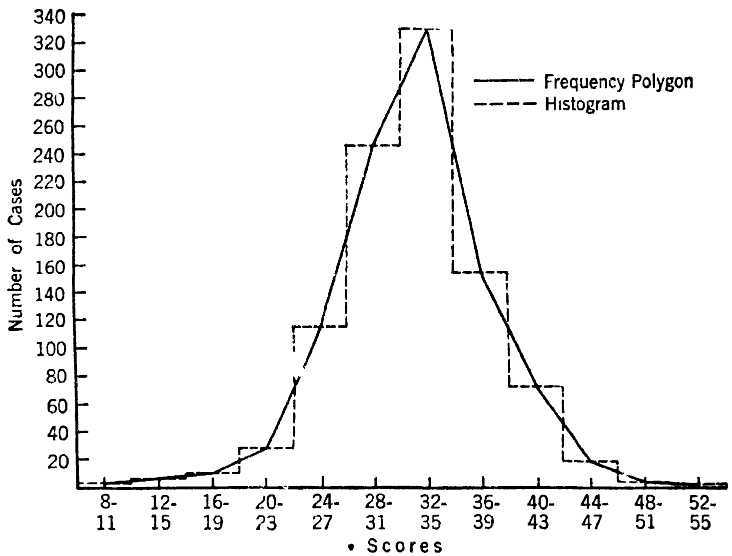


Fig. 4. Distribution Curves: Frequency Polygon and Histogram. (Data from Table 1.)

The facts brought out by a frequency distribution can be made more vivid if presented pictorially by means of a frequency graph. In Figure 4 are shown the data of Table 1 in graphic form. The base line or horizontal axis represents the scores; the vertical axis shows the frequency or number of cases falling within each class-interval.

The graph has been plotted in two ways, both being about equally common. One graph is a *frequency polygon,* in which the number of individuals within each interval is indicated by a point, centrally located in respect to the class-interval; the successive points are then joined by straight lines. The other graph is obtained by erecting a column or rectangle over each class-interval, the height of the column depending upon the number of cases in that interval. This type of graph is known as a *histogram.*

THE NORMAL CURVE

The reader will already have noticed certain characteristics of the distribution presented in Table 1 and Figure 4. The majority of cases cluster in the center of the range and as the extremes are approached there is a gradual and continuous tapering off. The curve shows no gaps or breaks; no clearly separated classes can be discerned. The curve is also bilaterally symmetrical, that is, if it should be divided by a vertical line through the center, the two halves so obtained would be nearly identical. This distribution curve resembles the bell-shaped "normal curve," the type most commonly found in the measurement of individual differences. The theoretically determined, ideal normal curve is illustrated by the graph reproduced in Figure 5.

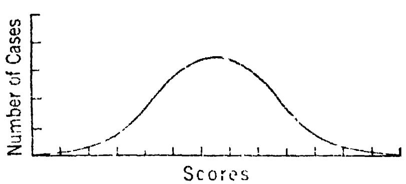


Fig. 5. Theoretical Normal Curve.

The concept of the normal curve is an old one in statistics. It first became familiar as the *normal probability curve.* The probability of the occurrence of an event is the expected relative frequency of occurrence of the given event in a very large, or infinite, number of observations. This probability is represented by a ratio or fraction, the numerator of which is the expected outcome, and the denominator the total possible outcomes. Thus the probability or chances that when two coins are tossed only heads will come up is ¼, or one out of four possible occurrences,[2] the probability of one head and one tail is ½; and that of two tails, ¼. If the number of coins is increased, say

[2] This follows from the fact that the only possible combinations of heads (H) and tails (T) which can occur when two coins fall are the following four: HH, HT, TH, TT. Just one of these four (HH) contains only heads.

to 100, so that the number of possible occurrences or combinations becomes very large, we can still determine mathematically the chances of any one combination, such as all heads or twenty heads and eighty tails, occurring. These probabilities, or expected frequencies of occurrence, can be plotted graphically by the same method outlined above for plotting scores. The curve obtained when the number of coins is very large will be the bell-shaped normal probability curve. In Figure 6 are shown the theoretical and obtained frequencies for 12

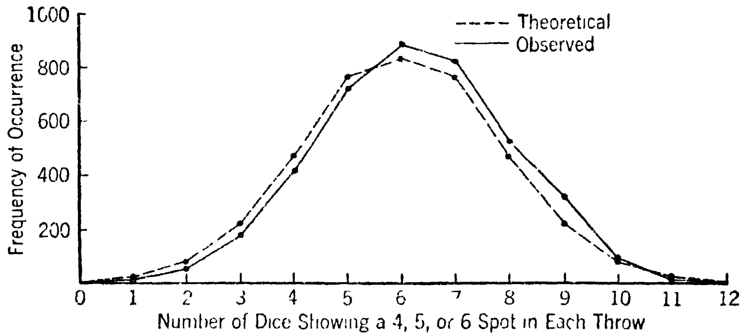


Fig. 6. Theoretical and Observed Distributions of Results in 4096 Throws of 12 Dice. (Data from Yule and Kendall, 35, p. 424.)

dice thrown 4096 times. In each throw, the number of dice showing a 4, 5, or 6 spot uppermost was determined. This number could, of course, vary from zero to 12, the total number of dice thrown. The graph shows the relative frequency of each combination in the total 4096 throws. It will be noted that there is a very close agreement between the theoretical and obtained curves.

The results obtained by tossing coins or throwing dice are said to depend upon "chance." By this is meant that the outcome is determined by a large number of similar, equal, and independent factors. The height from which a coin or die is thrown, its weight and size, the twist of the hand employed, and many similar conditions determine which particular face will fall uppermost. Likewise, a person's height, or weight, or performance on an intelligence test can be regarded as depending upon a variety of independent factors, each having about equal influence upon the result. Thus it has been suggested that the operation of chance is responsible for the distribution of human traits according to the normal frequency curve. It does not

follow, however, that if a characteristic is normally distributed, it is necessarily the result of "chance factors" as defined above.

The normal curve also appears in a different situation as the *curve of error* When repeated measurements are made, the results will not be identical on successive occasions Such fluctuations, or "errors," are present to a greater or lesser degree in all types of measurement The length of a table as measured by a meterstick, the speed of a simple movement, or the æsthetic appeal of a work of art will not remain the same on repeated observations If a very large number of observations of the same object or phenomenon are made, and the results found on successive occasions are plotted in a frequency graph, a normal curve will be obtained The errors of observation or measurement which produce the variation are themselves the result of chance factors, and hence the curve of error, like the distribution curve, will approximate the normal probability curve

OTHER TYPES OF DISTRIBUTION CURVES AND WHAT THEY MEAN

The implications of the normal distribution curve for a psychology of individual differences can be realized more vividly by contrasting this form of distribution with other possible types The distributions chosen in particular for this comparison are those implied by certain common theories and beliefs in regard to individual differences They are also occasionally found with actual test results because of the use of faulty techniques or the operation of special factors

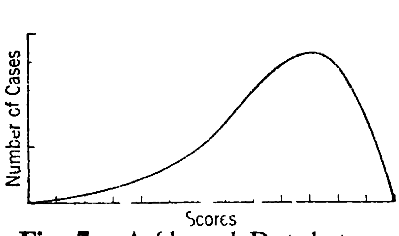


Fig. 7. A Skewed Distribution

A *skewed distribution* is one in which the peak or "mode" of the curve is displaced to either side of the center Such a distribution lacks the bilateral symmetry of the normal curve In Figure 7 will be found an illustration of a skewed curve, with a piling up of scores at the upper end of the distribution Such a distribution is implicit in the popular conception of many character traits Thus the majority of people are considered "honest" and are piled up at one extreme of the scale; from this point, the number of cases is believed to de-

crease steadily as the opposite extreme is approached. As will be illustrated in a later section of the present chapter, this type of distribution is not ordinarily found when adequate measures of character traits are used, i.e., measures which are capable of differentiating degrees of response.

In a number of behavior characteristics indicative of social conformity, a type of distribution known as the *J-curve* is often found. This curve, named after its resemblance to the letter J, is in reality a highly skewed curve, with the majority of people falling at that end which represents complete or nearly complete conformity. A favorite illustration of such J-curves is found in the reactions of motorists or pedestrians to various traffic regulations, such as stopping for traffic lights, stopping at intersections, or driving within the proper traffic lane. An example of such a curve is reproduced in Figure 15. Other illustrations of "conforming behavior" to which the J-curve has been applied include observations of religious practices, such as time of arrival at services, participation in group singing, amount of kneeling, and the like.

A type of distribution not so frequently found as the skewed curve but nevertheless assumed in certain common practices is the *rectangular distribution*, illustrated in Figure 8. If individual differences were distributed in this manner, it would mean that there were as many geniuses and idiots as mediocre people, as many men whose height is 6 feet 6 inches as those whose height is 5 feet 8 inches. It is interesting to speculate on the effect which such a situation would have on our sense of values. Our thinking is so permeated with the knowledge that extreme degrees of a trait are relatively infrequent, that it is difficult even to conceive of a world in which extremity did not imply rarity.

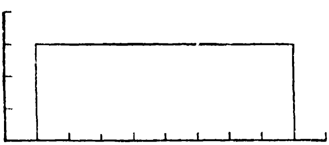


Fig. 8. A Rectangular Distribution.

The assumption of a rectangular distribution of traits is implicit in certain common misuses of percentile scores. In the percentile system of scoring, it will be recalled, the subject's standing on any test is expressed in terms of the percentage of people in a given group whose scores he excels. When comparing individuals who receive, let us say, percentile scores of 90, 80, 60, and 50, we must

bear in mind that the difference in ability between the first two cases is greater than that between the last two, although in both pairs the difference is 10 percentile points. In order to include the 10% of the cases which fall between the 90th and 80th percentiles, we must cover a much longer distance on the base line of the normal curve than is necessary in going from the 50th to the 60th percentiles. This results from the greater clustering of individuals near the center of the curve, and the relatively small number of cases at the extremes. *Only if the trait distribution were rectangular would successive percentile scores represent equal units of ability.* This does not mean that percentile scores are of no value. Like mental ages, they furnish a simple and vivid means of expressing the subject's standing on a test. Such devices do not, however, furnish an equal unit scale of ability. Neither percentiles nor mental ages, for example, lend themselves to averaging or to similar arithmetic operations, because of such inequality of units.

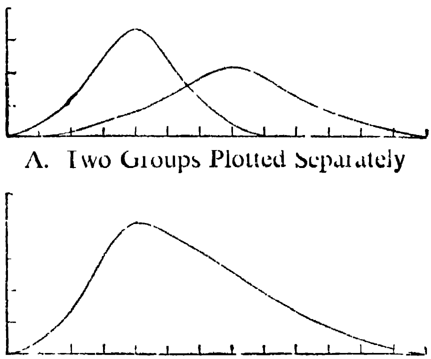
Lastly, special mention should be made of the *multimodal distribution* because of the prominent part it plays in so-called type theories. A multimodal curve is one having more than one mode or peak. Instead of a single clustering of individuals in the center as in the normal curve, or at either extreme as in a skewed curve, the clustering occurs at several points. The peaks may be equally large, or there may be a major peak and one or more minor ones. The most popular variety seems to be the bimodal curve, with two approximately equal peaks. All the common schemes of classification which place individuals into distinct categories presuppose some form of multimodal distribution. The division of men into the genius, the normal, and the feebleminded, the sane and the insane, the sociable and the unsociable, all rest upon a tacit assumption that "most people" can be classified clearly into one of these groups, with possibly a few intermediate doubtful cases. It is interesting to note that these distinctions are much less common in the realm of physical traits, where continuity of variation is more apparent to the naked eye.

CONDITIONS WHICH AFFECT THE SHAPE OF THE DISTRIBUTION CURVE

Distribution curves which deviate significantly from normality and which exhibit one or more of the properties discussed in the preceding

section occur from time to time because of the operation of special factors. A consideration of some of the most common of these factors is essential for the proper interpretation of frequency curves. The conditions which may influence the shape of the distribution curve include peculiarities of *sampling*, inadequacies of the *tests* or other measures employed, and factors which operate directly upon the distribution of the *behavior* itself. Among the last-named type of factors are pathological conditions and socially imposed constraints. In the following sections we shall consider each of these various conditions in turn.

Sampling. It would, of course, be possible to obtain any conceivable type of distribution by deliberately choosing subjects to fit the pattern. There would obviously be no object to such a procedure.



A. Two Groups Plotted Separately

B. Two Groups Combined

Fig. 9. Skewness Resulting from the Combination of Groups with Different Means and Variabilities.

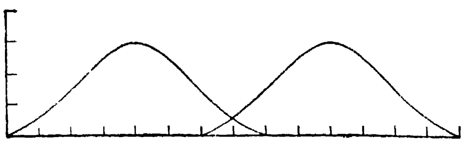
Similar variations may, however, occur through the operation of selective factors which may have been overlooked by the investigator. Whenever a curve deviates significantly from normality, the adequacy of the sampling ought therefore to be examined.

Skewness may result, for example, from the inclusion within a single distribution of two normally distributed groups which differ pronouncedly in both average and variability. This effect is illustrated in Figure 9. In Graph A are given the separate distribution curves of the two groups, one of which has a lower average as well as a narrower scatter of scores than the other. Graph B shows the definitely skewed

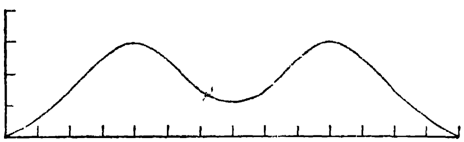
curve which is obtained when both groups are combined and plotted as one distribution.

A multimodal curve can also be obtained if the sampling tested is not chosen at random from the general population, but consists of individuals selected from widely differing levels and combined into a single group. A group consisting of 5-year-olds and 10-year-olds, for example, would present a definitely bimodal distribution in intelligence test scores, as well as in height, weight, and many other characteristics. Were the intervening age groups from 6 to 9 to be included in this sampling, the distribution would take on the appearance of the normal bell-shaped curve.

The production of a bimodal distribution by combining two curves of widely separated groups is illustrated in Figure 10. It will be



A. Two Groups Plotted Separately.



B. Two Groups Combined.

Fig. 10. Bimodality Resulting from the Combination of Two Groups with Widely Varying Means.

noted that the overlapping between the two groups is very slight. When the overlapping is large, as in the case of adjacent age groups, the resulting combined curve will be normal and unimodal. An example of a bimodal curve plotted with actual scores is presented in Figure 11. The two distributions which are combined in this curve consist of the Army Alpha scores obtained by two groups in the United States Army during World War I. The lower group includes 2773 native-born white soldiers who had reached no higher than the fourth elementary grade when they left school; the upper group consists of 3954 officers who had had four years of college work. The combined curve exhibits the definite bimodality which would be expected.

Other peculiarities which may result from sampling include exces-

sive flatness of the distribution curve (approximating a rectangular distribution), or its reverse, excessive peakedness. The latter might occur, for example, if the sampling is exceptionally homogeneous. Finally, it should be noted that an unlimited number of minor irregu-

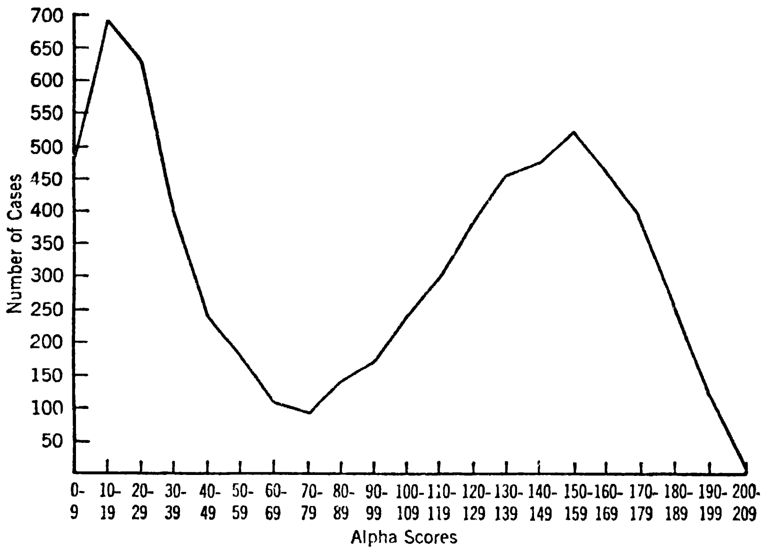


Fig. 11. A Bimodal Distribution Obtained by Combining Extreme Groups: Alpha Scores of 2773 Soldiers with 4th Grade Education and 3954 Officers with 4 Years of College. (Data from Yerkes, 34, pp. 773, 777.)

larities and variations in distribution curves may occur through the use of small groups. Curves plotted from a small number of cases usually present an uneven, jagged appearance, since individual exceptions loom relatively large. In general, the larger the sampling, the "smoother" will be the distribution curve.

Inadequacy of the Testing Range. If the range of difficulty covered by the test items is restricted at the upper or lower levels, a skewed curve may be artificially produced. Such a distribution will be obtained when any test is given to a group for which it is not suited. Thus if the National Intelligence Test, which is adapted to grades 3 to 8, were administered to a college class, the large majority of subjects would score very near the maximum, and the number of cases would decrease rapidly toward the lower scores. Similarly, if

one of the many tests constructed for use on college freshmen were given to elementary school children, there would be a marked piling of scores near the zero end of the scale, and the distribution would be equally asymmetrical.

Obviously these data could not be taken to mean that intelligence is not normally distributed among school children or college students. Such skewed distributions result from the fact that the difficulty range of the test does not extend far enough in the upper or lower direction. In the one case, all of those subjects who have more than a certain

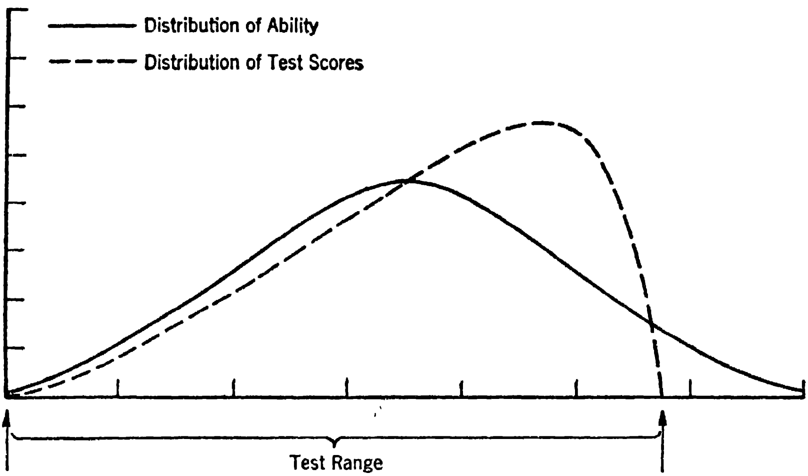


Fig. 12. The Effect of Restricted Testing Range upon the Form of the Distribution Curve.

minimum of the ability tested will make a perfect or nearly perfect score, whereas if the test had included more difficult items, these subjects would have scattered over a wide range. This is illustrated in Figure 12, the solid line showing the actual distribution of ability in the group, and the broken line the curve which would result from the use of a test with a low "ceiling." In a similar manner, a piling up of zero or very low scores will occur when the test is too difficult for the group. In choosing a test for a given group, therefore, care must be taken to insure that the subjects have sufficient leeway at both ends of the scale. The highest and lowest scores obtained should be a considerable distance from zero and perfect scores, respectively.

Inequality of Test Units. It can readily be demonstrated that inequality of units in the measuring instrument can distort a frequency distribution in various ways. A good illustration is furnished by data recently collected on visual acuity by means of two tests (27). The frequency distributions of the same group of 226 persons on each of these tests are shown in Parts A and B of Figure 13, respectively. Graph A is a sharply peaked and skewed curve obtained with the familiar Snellen chart, in which the subject's visual acuity index is

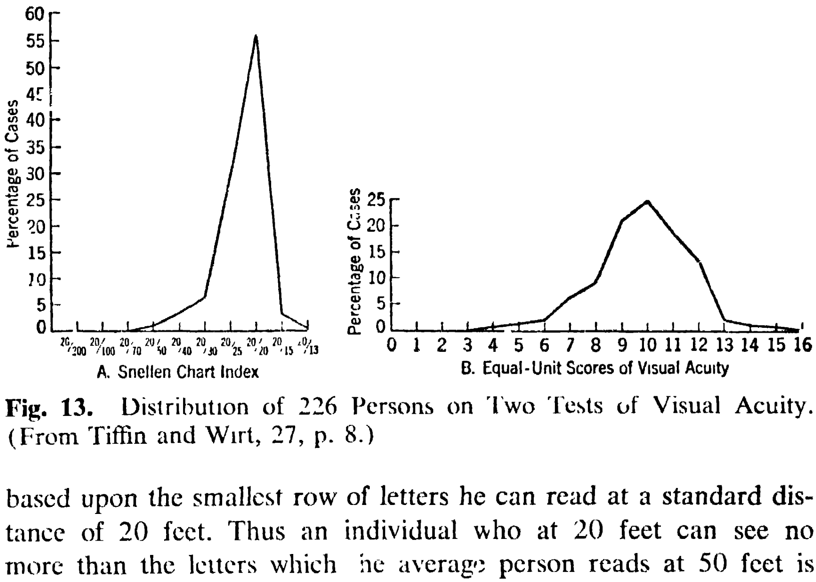


Fig. 13. Distribution of 226 Persons on Two Tests of Visual Acuity. (From Tiffin and Wirt, 27, p. 8.)

based upon the smallest row of letters he can read at a standard distance of 20 feet. Thus an individual who at 20 feet can see no more than the letters which the average person reads at 50 feet is said to have 20/50 vision. Normal vision obviously corresponds to a 20/20 index. An index such as 20/15 indicates better-than-average vision. Because of the particular choice of letter sizes in this test, not all acuity levels are sampled to an equal degree, the poorer acuity levels being represented by more items than the average or superior acuity levels. In other words, the differences in difficulty level between successive rows of letters are not equal; there are larger "gaps" in difficulty level in the center and upper portions of the acuity scale than in the lower portion.

This inequality of units can be illustrated by a comparison of the items or units on the Snellen chart with those on an equal-unit scale of visual acuity, as shown below (27, p. 9):

Acuity Scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Snellen Chart	20/200	20/100	20/70	20/50	20/40		20/30	20/25		20/20			20/15		20/13

The distribution of the scores of the same group of 226 persons on the equal-unit acuity test is given in Part B of Figure 13. It will be noted that this graph approximates a normal curve much more closely than does the distribution of unequal-unit scores.

TABLE 2 *Artificial Bimodality Resulting from Inequality of Units*

Equal Unit Scores	*Test Scores*	*Number of Cases*
15	1	1
16	2	3
17	3	7
18	4	12 } (29)
19		17 }
20	5	20
21	6	17 } (29)
22		12 }
23	7	7
24	8	3
25	9	1
		100

Other types of variation from normality of distribution may also result from inequality of test units. The artificial production of bimodality may be demonstrated by the highly simplified hypothetical example given in Table 2. Let us assume that the entries in column 1 represent equal units of a given ability and those in column 2 the corresponding scores on a nine-item test designed to measure this ability. We may further assume for simplicity that the nine test items are so steeply graded in difficulty that no subject can succeed with any one item if he has failed any previous item. In such a case, the total scores in column 2 will correspond exactly to the most difficult item which the subject is able to complete. It will be noted that certain ability levels (column 1) are not represented by test items (column 2). Thus there are no items to correspond to ability levels 19 or 22. The third column gives the distribution of 100 subjects in the ability under consideration. Obviously the 17 persons falling at ability level 19 do not have enough ability to succeed with item 5, which requires ability level 20; they will therefore have to stop with item 4 and thus augment

the group of 12 persons who have just barely enough ability to complete item 4. The same will occur in the case of the 12 persons who fall on ability level 22.

The distribution curves of the equal-unit scores and the test scores, respectively, are given in Figures 14A and 14B. It will be noted that,

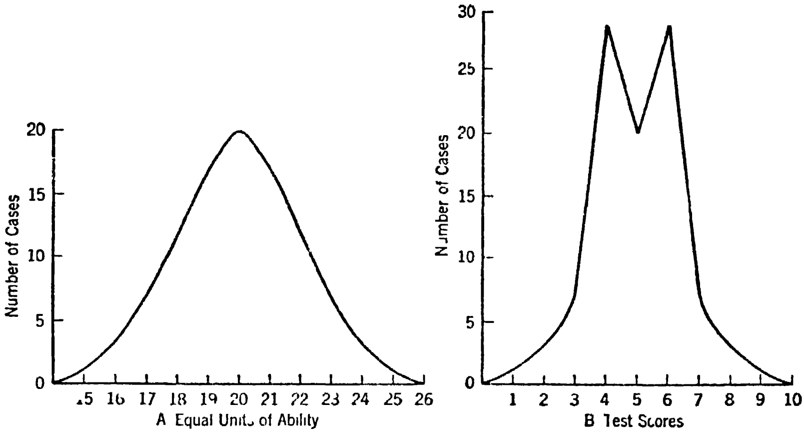


Fig. 14. Hypothetical Distributions of 100 Persons on an Equal-Unit and an Unequal-Unit Measure of the Same Ability.

although the former is practically a normal curve, a distinct bimodality is introduced in the latter by the inequality of units. It is well to bear this effect in mind when considering results obtained with tests of certain personality characteristics, such as introversion-extroversion. Being defined in bipolar terms, such traits may have been sampled more thoroughly in their extreme manifestations, while their intermediate degrees may have insufficient coverage. Such a test would thus have poorer discriminative value and larger gaps between units in the center of the range than at the extremes, as in our hypothetical illustration of Table 2. As a result, a slight bimodality could easily occur simply from the peculiarities of the measuring instrument, regardless of the distribution of the behavior itself.

Pathological Conditions. Deviations from normality of distribution may result from conditions which affect the development of the behavior itself, rather than from characteristics of the test or of the sampling. An example of such an effect is to be found in the distribution of IQ's. When the *total population* is considered, the distribution

shows an excess of extremely low IQ's over what would be expected in a normal curve. In one extensive survey conducted in England, for example, the proportion of cases with IQ's below 45 was about 18 times as great as would be expected in a normal distribution with the obtained mean and SD [3] (20).

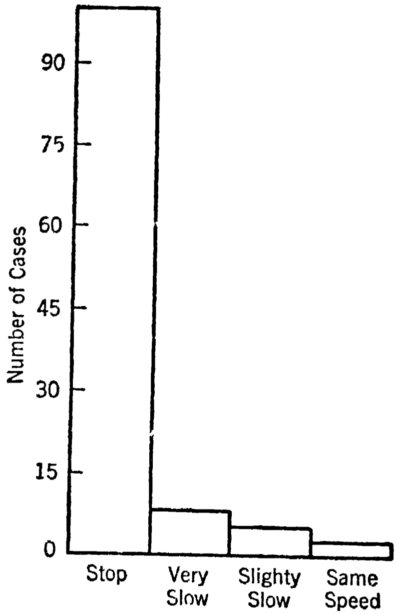
The most plausible explanation for such a deviation from normality would seem to be that secondary factors, such as disease or pathological conditions, increase the relative proportion of feebleminded persons (cf., e.g., 14). It will be recalled that a normal distribution will be obtained if the variable being measured is the composite result of a very large number of independent and equally weighted factors. Considering the extremely large number of both hereditary and environmental factors which contribute to the development of intelligence in the general population, it is reasonable to expect IQ's to distribute themselves in accordance with the normal curve. If, however, any factors should operate with disproportionate weight, then the effect on the curve would be equivalent to the use of loaded dice in disturbing "chance" results. Pathological conditions, which may lower the IQ but can never raise it, may be regarded as such "loading" influences. It should be noted that the data concerning the lower end of the distribution of IQ's, as well as the interpretation of such data, are still highly tentative. They are here cited merely as an illustration of the possible effect of pathological conditions upon distribution curves.

Socially Imposed Constraints. Another factor which may "load the dice" and alter the distribution of behavior characteristics is to be found in socially imposed barriers. Such conditions often produce the highly skewed J-curve described in an earlier section. The effect of social constraints upon the form of the distribution can be illustrated by the behavior of motorists. At an ordinary intersection with no traffic signal, the behavior of drivers will probably follow the normal curve, the majority exhibiting a moderate amount of caution, very few coming to a full stop, and equally few continuing at the same rate of speed with no observation of oncoming traffic. If, now, red signal lights and a policeman are installed at the intersection, these external constraints will pull the distribution into a J-curve. Figure 15 shows the distribution of the responses of 102 motorists at an intersection with

[3] Cf. fuller report of this investigation on pp. 81-82.

no cross traffic approaching, but with red signal lights and a traffic officer.[4] It will be noted that over 90% came to a full stop. Of the remaining small per cent, a few slowed down markedly, still fewer slowed down slightly, and a very small number continued at the same speed (1).

It should be noted that the location of the peak depends upon the point in the scale at which the socially imposed behavior falls. The extreme or true J-curve is not necessarily obtained in all situations involving social conformity. Thus the degree to which urban adults in America partake of alcoholic beverages would probably show a peak, not at either extreme, but at an intermediate point corresponding to "moderate social drinking." This point probably represents maximum conformity to the practices of the group, but it does not represent either a maximum or a minimum in terms of drinking behavior. It is not the J-curve itself that is important, but rather the fact that variations in the distribution curve may be introduced by social conformity. The J-curve is only a special instance of the effects of this type of "loading" factor.[5]



Response at an Intersection with no Cross-Traffic but with Red Signal Light and Traffic Officer

Fig. 15. J-Curve of Motorists' Behavior. (From F. H. Allport, 1, p. 144.)

[4] This curve resembles the letter L more than it does the letter J, but it has become conventional to refer to all such highly skewed curves as J-curves, regardless of whether the peak is at the extreme right or extreme left. The direction of the scale could, of course, be arbitrarily reversed in all such cases, so that the peak would be at the right.

[5] It has been suggested by some writers (e.g., 2, pp. 332–337) that the normal curve may be regarded as two J-curves "back-to-back," so to speak (JL). The normal curve may, of course, be broken up in an infinite number of ways and conceived as a composite of any number of arbitrarily separated parts. This does not, however, in any way alter the characteristics of the total distribution, nor the mathematical properties of the curve.

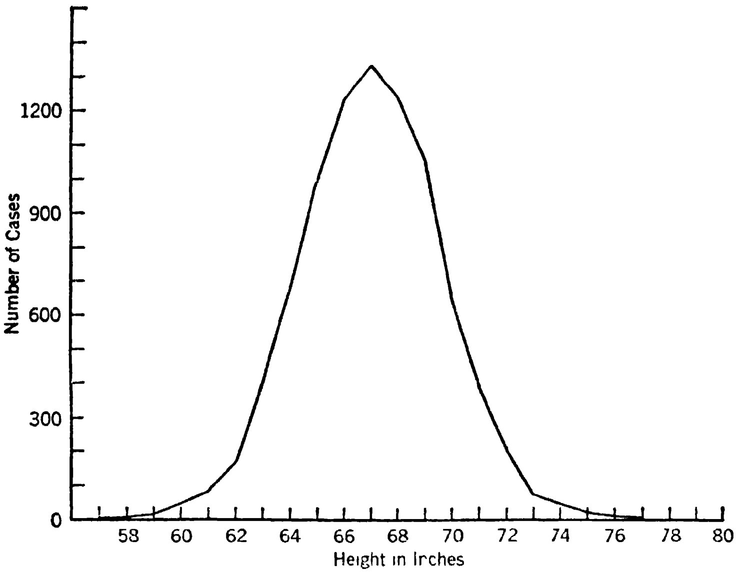


Fig. 16. Distribution of Height for 8585 Adult English-Born Men. (From Yule and Kendall, 35, p 95)

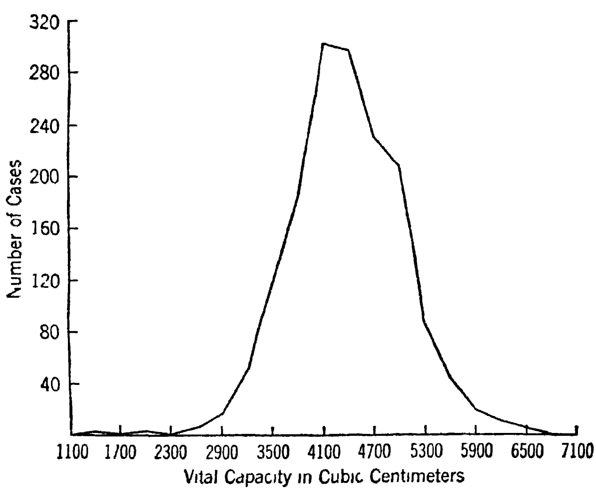


Fig. 17. Vital Capacity of 1633 Male College Students. (From Harris et al., 8, p 94)

SOME TYPICAL DISTRIBUTIONS

In Figures 16 to 31 will be found examples of distribution curves obtained for a wide variety of human characteristics. These distributions were chosen principally because they were based on large, representative samples, most of them including 1000 or more cases. A few curves plotted from smaller groups have been included to illustrate the distribution of physiological and of certain personality characteristics, since in these areas data on large groups are relatively scarce.

An example of the distribution of a purely structural trait is furnished in Figure 16, which shows the *height* in inches of 6194 English-born men. It will be seen that the graph approximates the mathematical normal curve to a remarkably close degree. Figure 17 presents the frequency curve of a more functional, physiological trait, *vital capacity*. This is the total volume of air, measured in cubic centimeters, that can be expelled from the lungs after a maximal inspiration. The measurements from which the curve is plotted were made on 1633 male college students. The general correspondence to the normal curve is again apparent.

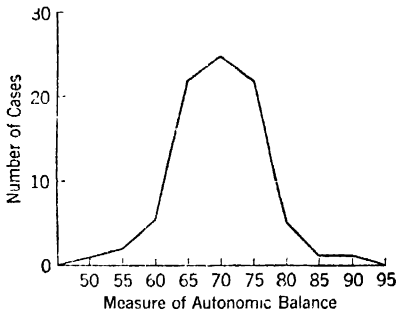


Fig. 18. Distribution of Mean Estimates of Autonomic Balance for 87 Children between the Ages of 6 and 12. (From Wenger and Ellington, 33, p. 252.)

Figures 18 and 19 are concerned with physiological measures which are believed to have some relationship to emotional and personality characteristics. The first shows the distribution of 87 children in a composite measure of *"autonomic balance."* High scores in this measure indicate a functional predominance of the parasympathetic division of the autonomic nervous system; low scores, a functional predominance of the sympathetic division. To psychologists, the autonomic nervous system has been of special interest because of its role in emotional behavior. The distributions of 74 children in two different indices of *muscular tension* are shown in Figure 19.

The two graphs reproduced in Figures 20 and 21 illustrate the distribution of performance on *sensori-motor* and *simple learning* tests.

Reference may also be made in this connection to the data reported previously in Table 1 and Figure 4. All three sets of measures were obtained on the same group of 1000 college students. The tests whose distributions have been reproduced include cancellation, Pyle symbol-digit, and a nonsense-syllable "vocabulary" test. In the first, the score is the total number of A's in a page of pied type cancelled in one minute. This is generally regarded as a simple test of attention and perception, although speed and control of movement are also involved. The symbol-digit test is a simple learning test of the code substitution variety. The vocabulary test is a more difficult learning test, also

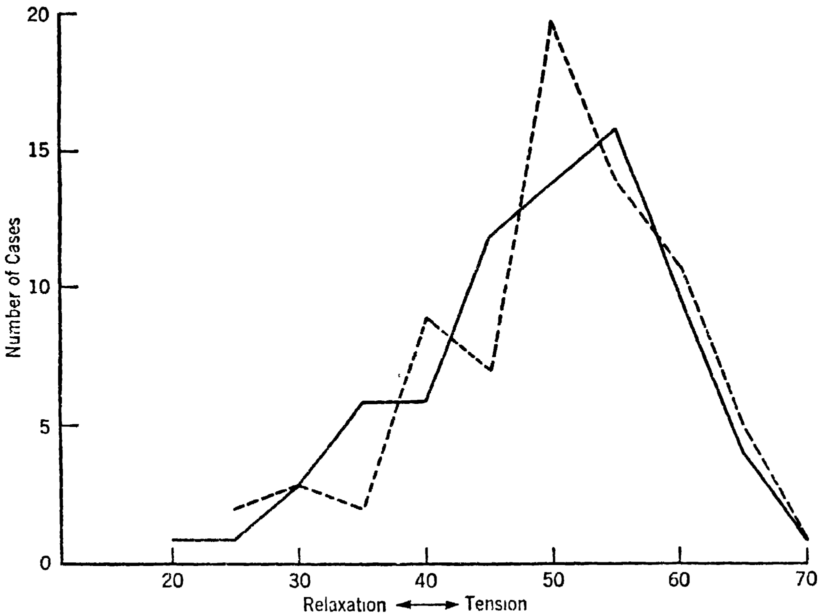


Fig. 19. Distribution of Two Measures of Muscular Tension for 74 Children between the Ages of 6 and 12. (From Wenger, 32, p. 222.)

employing a code, which in this case consists of paired nonsense syllables. The distributions of all three tests fall within the expected values of the theoretical normal curve.[6]

Typical results obtained with *intelligence tests* administered to large samplings are presented in Figures 22 to 26. Figure 22 gives the dis-

[6] Mathematical tests of normality were applied to these curves (cf. Anastasi, 3).

tribution of the IQ's of 2904 children between the ages of 2 and 18 on the 1937 revision of the Stanford-Binet. Reference to the graph will show that the largest per cent of cases received IQ's in the middlemost class-interval, from 95 to 104. The per cent tapers off gradually until only a small fraction of 1% is found with IQ's between 35 and 44, and between 165 and 174. Institutionalized feebleminded subjects were not included in this distribution, the sampling also being restricted in certain other ways. Thus the group consisted entirely of American-born white subjects, with a somewhat greater proportion of urban

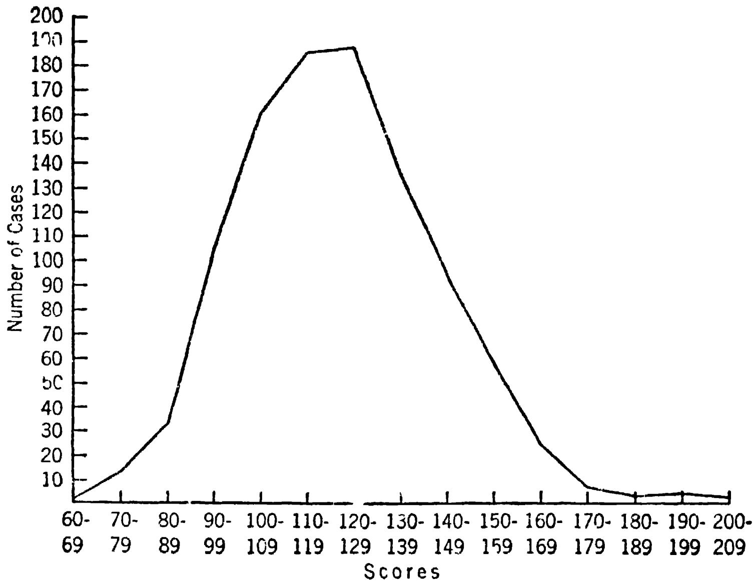


Fig. 20. Number of A's Cancelled in One Minute by 1000 College Students. (From Anastasi, 3. p. 32.)

residents than is found in the total population of the country. The major portion of the sampling was composed of elementary school children, an effort having been made to secure groups at the younger and older ages which were roughly comparable to the elementary school population. It might be noted that the range of IQ's for the total population, as determined from the data of various investigators, actually extends from nearly zero to slightly over 200.

Distributions of scores on group tests, obtained with children as

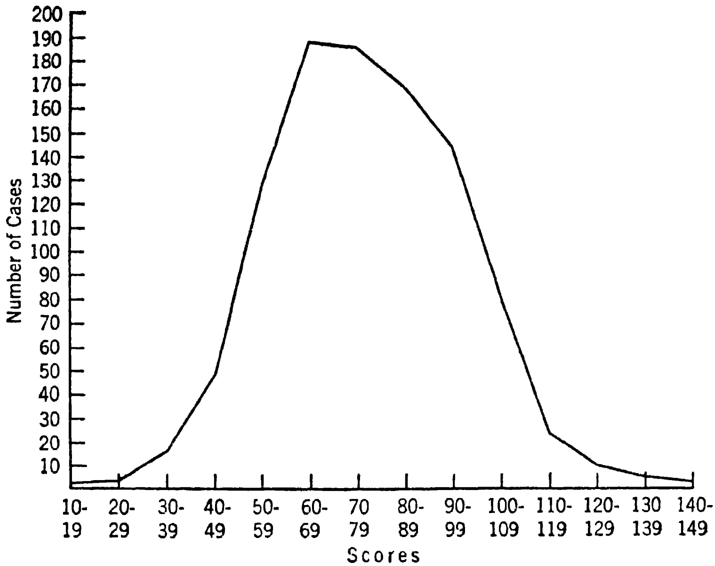


Fig. 21. Scores of 1000 College Students on a Symbol-Digit Code-Learning Test (From Anastasi, 3, p 34)

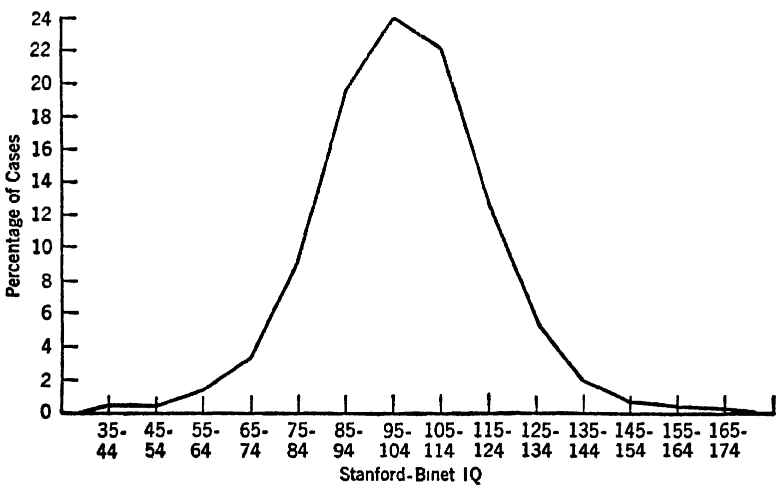


Fig. 22. Stanford-Binet IQ's of 2904 Unselected Children between the Ages of 2 and 18. (From Terman and Merrill, 24, p. 37)

well as adults, are illustrated in the next two graphs. Figure 23 shows the percentage distribution of 5952 sixth grade school children on the Advanced Otis Examination, a widely used group test of general intelligence.

The distribution of the AGCT scores obtained by 9,339,289 men during World War II is reproduced in Figure 24. This distribution exhibits two noteworthy deviations from the general form of the normal curve (23). The most conspicuous deviation is the sudden piling up of individuals as the standard score of 40 is approached, a score which is close to the actual zero point of the test. The AGCT is unsuitable for measuring the abilities of persons having less than the equivalent of a fourth grade education. Consequently, when this test is administered to an unselected sampling, a piling up at or near a raw score of zero will occur. Many of the individuals who fell into this category were illiterate, and for most of them the zero score merely indicated that they should be reexamined with a nonverbal test. The broken line in Figure 24 shows the extrapolated end of the distribution which would probably have been obtained if the test had had a much lower zero point. A further characteristic to be noted in the curve is the small bulge between the scores of 60 and 80. One explanation which has been offered for this bulge is that a considerable proportion of the population have little interest in continuing such academic activities as reading and arithmetic after leaving school and they therefore allow these skills to retrogress (23). These individuals would thus make a poorer showing on such a test than they would have made earlier.

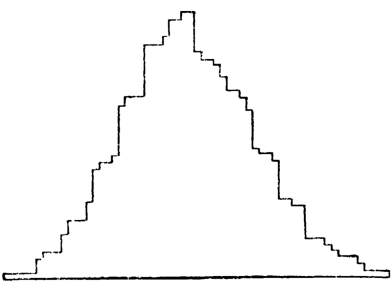


Fig. 23. Percentage Distribution of 5952 Sixth Grade School Children on the Otis Advanced Examination. (From Thorndike et al., 26, p. 5[illegible])

Of special interest are a few intelligence test surveys conducted on complete or nearly complete populations of children. In one of these (cf. 19, 20), the population chosen consisted of all children born between September 1, 1921 and August 31, 1925 and living within the boundaries of the city of Bath, England, on July 27, 1934. The Advanced Otis Examination (Form A) was given to all except the

defective children, who were subsequently tested with the 1916 revision of the Stanford-Binet. All children falling below a certain score on the Otis were also retested with the Stanford-Binet (cf. 20). An unusually close approximation to the desired population was achieved, the number actually tested being 3361 out of a total of 3398 cases which fell within the specifications given above. The distribution of intelligence for the group as a whole did not deviate significantly from the normal curve. The fit was good for all portions of the obtained

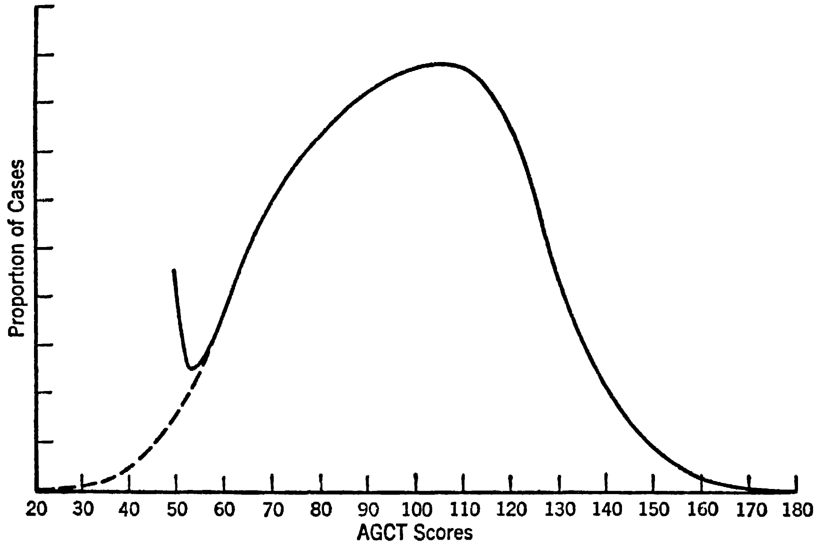


Fig. 24. Scores of 9⅓ Million Men on the Army General Classification Test. (From Personnel Research Section, A.G.O., 23, p. 415.)

distribution with the exception of the lowest 1% of the group, whose IQ's were 63.4 or below. The number of children with lower IQ's was in excess of the expected proportion, although the deviation was not significant until about the level of IQ 45. Below this level the excess was marked.[7]

Probably the most ambitious sampling project undertaken to date was the testing of all children who had been born in Scotland in 1921, with the exception of the blind and the deaf (22). A specially de-

[7] Cf. the citation and discussion of these same findings on p. 74.

signed 45-minute group test consisting of two pages of pictorial and five pages of verbal items was employed, together with a preliminary 10-minute practice test. All testing was done on June 1, 1932,[8] the children therefore ranging in age from 10½ to 11½ at the time of testing. A total of 87,498 children were tested, a sampling which the authors describe as complete except for a negligible number of children in certain private schools and a few who were absent through sickness or other causes. It will be noted, moreover, that the testing occurred at a time of the year when absences are at a minimum. Separate scores are reported on the verbal and pictorial parts of the test.

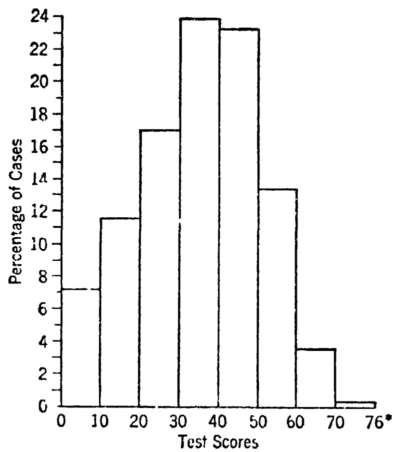


Fig. 25. Distribution of the Scores of 87,498 Scottish Children on a Verbal Group Test of Intelligence. (Data from Scottish Council for Research in Education, 22, p. 61.)

* The last class-interval does not cover 10 points, since the maximum score on the test was 76.

The percentage distribution of the verbal scores is given in Figure 25. Although on the whole this distribution shows a single clustering of scores at the center and a progressive decrease in frequency as the extremes are approached, a number of irregularities can be noted. Inequality of test units and inadequate coverage at the low end of the scale are strongly suggested by a consideration of the test itself. The fact that 7.2% of the cases fell in the class-interval 0–9 further indicates that the zero point of the test was probably set too high for the present population. With the inclusion of more easy items, these cases would very likely have distributed themselves over several class-intervals, below the present zero of the test. The distribution of the pictorial scores revealed the opposite effect, the test evidently being fairly well suited for the low-grade cases but too easy for the majority of the children. This distribution was highly skewed, with a marked piling up of cases at the upper end.

[8] Except in two areas where local circumstances necessitated testing on June 2 and 3, respectively.

What is undoubtedly the most nearly complete testing of an entire population is to be found in the second Scottish survey (17, 25), reported in 1939. The population chosen for this survey was considerably smaller, including only children born in Scotland on any one of four specified days in 1926 (Feb. 1, May 1, Aug. 1, and Nov. 1). A diligent and painstaking search to the remotest corners of Scotland finally yielded a complete sampling, with the loss of only one case. The group included a total of 443 boys and 430 girls, ranging in age from 8-11 to 11-9 at the end of the testing. All children were given the 1916 revision of the Stanford-Binet[9] together with eight performance tests selected from the Pintner-Paterson Performance Scale and other available series of a similar nature.

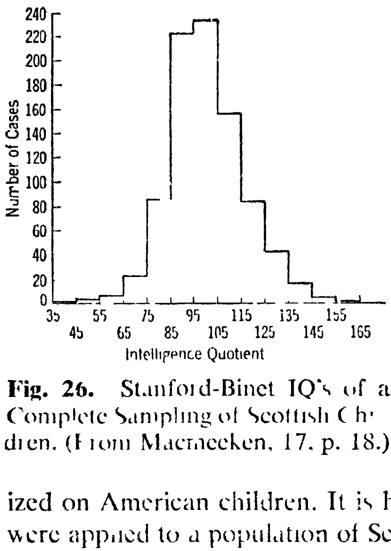


Fig. 26. Stanford-Binet IQ's of a Complete Sampling of Scottish Children. (From Macmeeken, 17, p. 18.)

The distribution of Stanford-Binet IQ's for the total group of 873 children is reproduced in Figure 26. It will be seen that, although again exhibiting the general form of the normal curve, this distribution deviates from the theoretical normal curve in a number of specific ways. The distributions of scores on the various performance tests also showed a number of irregularities. These results are not surprising when we consider that the Stanford-Binet as well as most of the performance tests used in this survey had been standardized on American children. It is highly probable that when such tests were applied to a population of Scottish children, the relative difficulty of units and the significance of raw scores were appreciably altered. These changes would in turn affect the shape of the distribution curves.

In the measurement of *personality* and *character*, testing techniques are still in a relatively crude and undeveloped stage. Many sources of error remain, so that one should scarcely expect to find perfect specimens of the normal distribution curve. Despite a more jagged appearance and many minor irregularities, however, the available distribu-

[9] The 1937 revision had not yet appeared when the testing was begun.

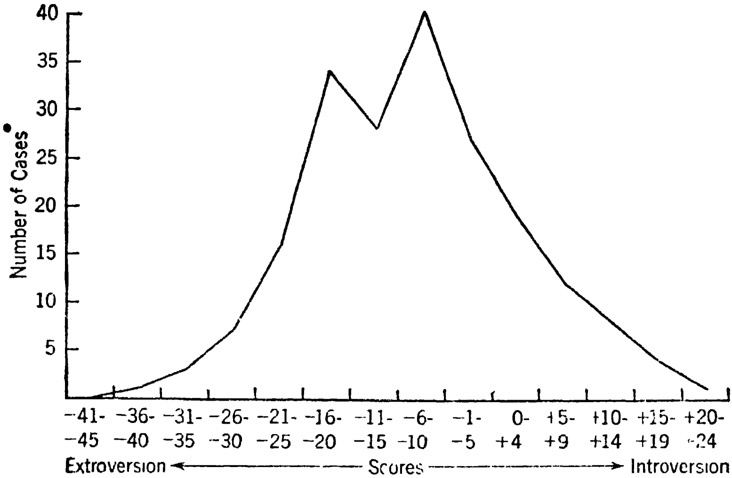


Fig. 27. Distribution of Introversion-Extroversion Scores of 200 College Students. (Data from Heidbreder, 12, p. 124.)

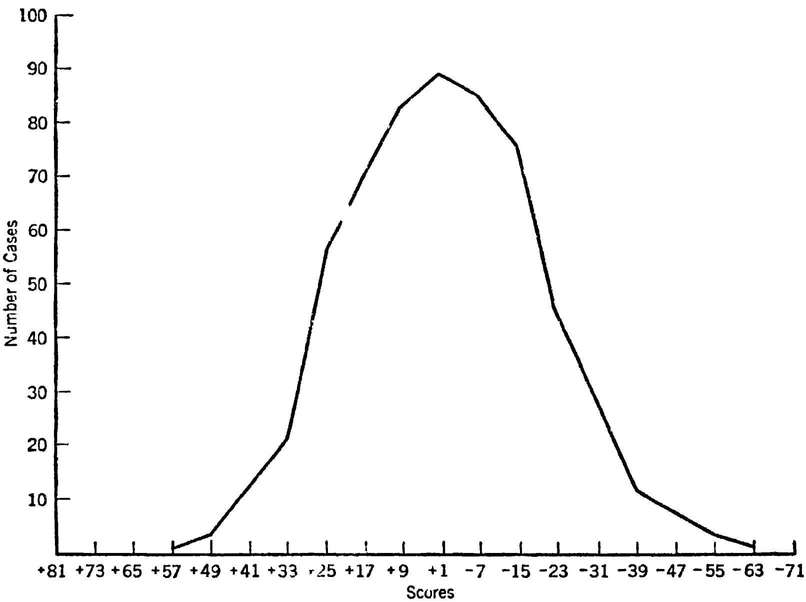


Fig. 28. Distribution of 600 College Women on the Allport Ascendance-Submission Test. (From Ruggles and Allport, 21, p. 520.)

tion curves exhibit quite generally the fundamental characteristics of the normal curve. Inspection of Figures 27 to 31 will make this apparent.

Figure 27 gives the distribution of total introversion-extroversion scores on a self-rating questionnaire administered to 200 college students. The positive scores correspond to the introvert end of the scale, the negative scores to the extrovert end. It will be readily seen that individuals do not cluster at opposite ends of the scale, as a clear-cut division into introverts and extroverts would imply. The greatest clustering occurs in the center, with a gradual dropping off as the extremes are approached.[10] Figure 28, showing the distribution of 600 college women on a test of ascendance-submission, closely approximates the normal curve.

Figures 29 to 31 are plotted with data taken from the studies of May and Hartshorne (9, 10, 11) on the measurement of character in school children. Figure 29 gives the distribution of "cheating ratios" for 2443 children. The cheating ratio indicates the number of times each child cheated relative to the number of opportunities offered. The obtained curve does not admit of a clear-cut division of the group into the "honest" and the "dishonest," or into those who cheat and those who do not cheat. A slight skewness is exhibited, with a tendency for scores to pile up at the "honest" end, but this may be caused by a limitation in the scale. The tests probably presented an insufficient number of situations in which cheating was made very easy or in which it involved a relatively minor "moral issue." This would cut the scale short at the lower end and produce an excess of zero or very low cheating scores.

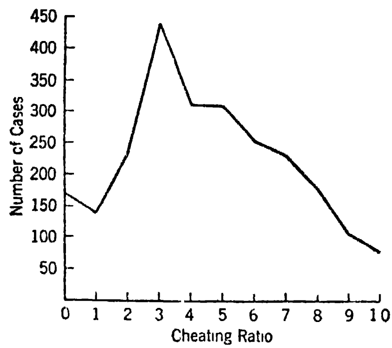


Fig. 29. Distribution of "Cheating Ratios" of 2443 School Children. (Data from Hartshorne and May, 9, p. 220.)

Figure 30, giving the distribution of combined scores on several

[10] The slight bimodality near the center of the scale might result from larger "gaps" between the items designed to sample intermediate degrees of this behavior, as discussed in an earlier section.

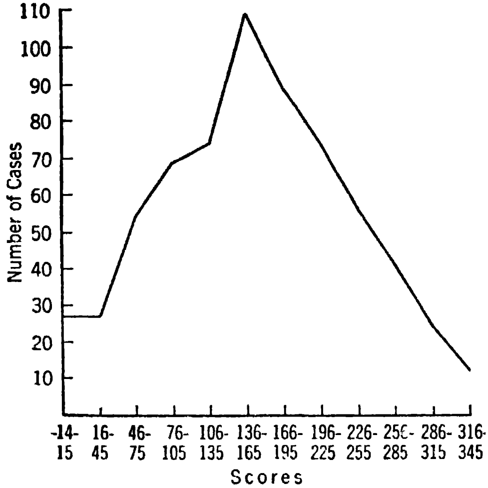


Fig. 30. Distribution of Persistence Scores among 656 School Children. (Unpubl. data from investigation of Hartshorne, May, and Maller, 10.)

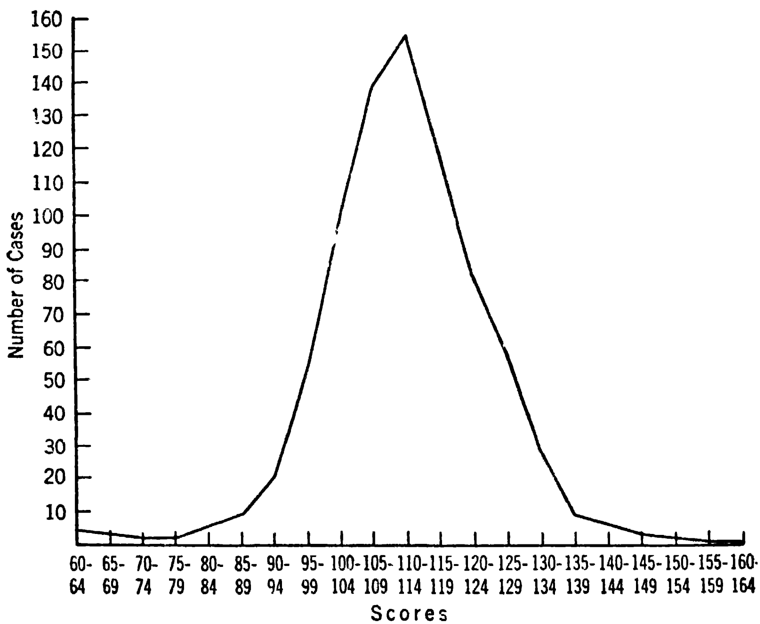


Fig. 31. Distribution of Cooperativeness among 801 School Children. (Unpubl. data from investigation of Hartshorne, May, and Maller, 10.)

tests of persistence, likewise exhibits the general characteristics of a single central peak and approximate bilateral symmetry A particularly close resemblance to the normal curve is to be found in the distribution of combined scores on several tests of service, or cooperativeness, presented in Figure 31

THE NORMAL CURVE AS A METHODOLOGICAL PROBLEM

As applied to psychological characteristics, the normal distribution curve should be regarded more as a methodological problem than as a factual observation Strictly speaking it is impossible to determine the actual distribution of a variable unless an equal-unit scale of measurement is employed The effects of inequality of units have been illustrated in an earlier section The only methods now available for obtaining equal units in a psychological test are however, based upon the assumption that the behavior under consideration is itself normally distributed Thus to ask what is "the" actual distribution of behavior constitutes at least for the present a meaningless question

A more significant inquiry however, concerns the specific conditions which determine the shape of the distribution curve in any particular situation In approaching this question we may begin with the fact that the distribution most likely to result when a characteristic depends upon a large number of independent and equally weighted influences is one resembling roughly the normal curve The reasonableness of this expectation for psychological characteristics is supported by the known complexity of their determination It is also relevant to note that nearly all physical traits, which are measured with equal unit scales do follow the normal curve

If then we begin with the expectation [11] that distribution curves will in general resemble the normal curve, any deviation from normality becomes a problem for investigation Such an approach to the form of the distribution should prove fruitful in revealing the operation of factors which merit study in their own right For example a significant deviation from normality may indicate that the test ceiling is too low, or that its zero point is too high, for the group being tested Similarly, some hitherto unsuspected selective factor operating in the sampling under investigation may now become apparent Finally, the shape of the obtained distribution may furnish a clue to an important

[11] This is not proposed as an assumption regarding the form of the distribution of abilities but is a promising starting point in the investigation of specific distributions

influence whose operation modifies the behavior itself in such a way as to alter the distribution curve. In other words, any significant deviation from normality should serve as a "signal" to alert the investigator to the need for further research.

It is certainly apparent that in the process of test construction the normal curve is now implicitly treated as a methodological concept, rather than as an empirically observed datum. Whenever a non-normal distribution is found in the standardization group, the immediate response is to set to work revising the test. Most tests have thus been deliberately adjusted so as to yield a distribution which approximates the normal curve in the population for which they were designed. Items are dropped or added, tests are shifted up or down in the scale, scoring "weights" of different responses are altered, and other similar adjustments are made until the desired approximation to normality is attained. To say, then, that a given distribution is normal may simply mean that the process of test standardization was meticulously executed. Conversely, to say that a given distribution is *not* normal may mean only that the construction of the test was crude, or that the test was applied to a group unlike the standardization population.[12]

THE MEASUREMENT OF VARIABILITY

One is tempted to compare the distributions of different traits in the effort to discover the relative variability of such traits. Do individuals differ more in physical or in psychological traits? Are they more alike in intellectual or in emotional characteristics? These and many similar questions have been raised repeatedly and answers have occasionally been offered.[13] It is probably correct to state as a general principle that individual differences will be larger in the more complex than in the simpler traits. Any characteristic which depends upon the simultaneous variation of a large number of factors will exhibit more marked differences than one which is determined by relatively few factors. An illustration from coin tossing will again prove serviceable.

[12] To argue that psychologists have been "biased" in favor of the normal curve and that non-normal distributions ought to be accepted (cf., e.g., 7) is just as meaningless as the insistence that the normal distribution of behavior has been empirically established. With the existing procedures of test construction, non-normal distributions are no more independent of the measuring instrument than are normal distributions.

[13] Cf., for example, the interesting although rather futile discussions by Wechsler (31) and by Ellis (5). The former treatment fails to come to grips with the fundamental difficulty presented by arbitrary test zeros, while the latter is vitiated by several instances of faulty reasoning and factual error

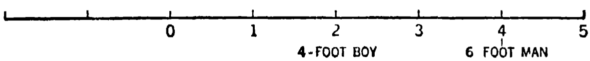
If two coins are employed, the number of possible combinations which may result is only four; if, however, the number of coins is increased to ten, the possible variations, or patterns of head-and-tail combinations, total 1024. A complex trait is one which is determined by a large number of factors or conditions, and hence it will be expected to exhibit a greater range of variation.

Apart from this rather obvious generalization, little can be said about the extent of individual variation in different traits. Upon close analysis, in fact, the question of the extent of variability itself appears to be ambiguous and quite meaningless. The first problem which confronts one when trying to compare human variability in separate traits is that of the measuring rod employed for the different traits, or the units in which the measurements are reported. That the particular scale employed affects the amount of variability found is easily demonstrated. If the height of buildings in one city is measured in feet and in another city in yards, the buildings in the former city will seem to vary among themselves three times as much as in the latter, even though the actual range in height may be identical in the two cities. Fortunately, feet can be translated into yards and vice versa. But this cannot be done with the units of psychological tests. The number of problems correctly solved on an arithmetic test cannot be transmuted into the same kind of units as words in an analogies test. The only solution offered for this difficulty is the use of measures of *relative variability*.

All indices of relative variability are ratios. One such measure, the coefficient of variation, is found by dividing the standard deviation (cf. Ch. 2) by the average of the distribution. Thus variability is expressed in relation to the average, the difference in units from one test to the other being automatically ruled out. For the same purpose, the ratio between the highest and lowest scores or the tenth highest and tenth lowest, or any other similar combination, is occasionally computed. Although in current use, all such measures are open to serious criticism. The difficulty arises from the fact that psychological scales do not measure the individual from a true or *"absolute zero"* of ability as a base. A zero score on the National Intelligence Test, for example, does not mean zero intelligence. This test begins at an arbitrary level corresponding to the ability of an average third grade school child. Consequently, anyone who fails to reach this level will receive a zero score on the test. If such an individual is given a test with a lower

"zero point," such as a first grade or preschool test, for example, his score will no longer be zero. A zero score on a psychological test is thus an arbitrary zero, which varies from test to test.

The custom of measuring from "absolute zero" in our physical scales is so general that it is difficult to conceive of the effects of using a scale that begins at an arbitrary zero point. Let us imagine a measuring stick on which height is measured, not from absolute zero or no height at all, but from some arbitrary point such as two feet. The following diagram illustrates the situation. Any object two feet or less in



height would register zero on this scale. If such a scale were to be employed only to measure the heights of individuals over five years of age, the arbitrary limit would perhaps not appear so absurd, since no one would be under two feet tall. This is in fact what has occurred in the construction of psychological tests. Since the AGCT, for example, was designed for adult men, it would have been wasteful, and from a practical standpoint impossible, to extend it down to the intellectual level of a newborn child.

To return to our yardstick with an arbitrary zero point at two feet, let us suppose that it has been used to measure the heights of a six-foot man and a four-foot boy. The man will measure four feet and the boy two feet, as has been indicated on the diagram. For many purposes, no error has been introduced in the data by the use of the artificial zero point. On any scale, the man is two feet taller than the boy. If, however, we express their respective heights as a ratio, we reach the conclusion that the man is twice as tall as the boy (4/2). This is not true of their actual heights from absolute zero, the man being only 1½ times as tall as the boy (6/4). The subtraction of a constant, two feet, from both heights has distorted the ratio.

Such is the effect of an arbitrary zero point on any value which involves the *division* of one measure by another. For this reason, ratio or other relative measures cannot be employed in comparisons among the large majority of psychological tests, which are not scaled from absolute zero.[14] Such measures would hold true only for the specific

[14] The only important exception to date is the CAVD Intelligence Examination, prepared by the Institute of Educational Research at Teachers College, Columbia University (cf. Thorndike, *et al* 26).

tests in the form in which they were employed; the addition or removal of a few easy items at the lower end of the scale would completely alter the relative variabilities. Obviously the values thus computed could not be regarded as very meaningful. We thus arrive at the conclusion that *with available psychological tests it is impossible to compare variability from one trait to another.*

Other difficulties also appear as the problem is inspected more closely. Does the question of the extent of variability refer to the whole human race? Which individuals, if any, shall be omitted in order to arrive at an estimate of human variability? Shall those who are regarded as definitely pathological and represent extreme deviations be excluded? If so, where should the line be drawn between a typical human group and an abnormal deviant? It seems reasonable, for example, to exclude from an estimate of the range of human variation in speed of movement one who has suffered an injury which renders his movements slow and halting. It is but a short step from this procedure to that which would exclude those incapacitated through disease. How, then, would this criterion operate in the case of a feebleminded person in whom no physical defect can be discovered? How far shall this process of eliminating extreme cases be carried?

A further question relates to the factors which are to be held constant in measuring the variability of any one trait. How homogeneous should the group be? The inclusion of children of different age levels would certainly increase the extent of variation in most traits. If only the range of individual differences within a fairly homogeneous population is desired, the difficulty of defining the required degree of homogeneity is encountered. Many traits are influenced by the social and economic level in which the individual finds himself. Should conditions of this sort also be held constant? Should differences in speed of performance be ruled out when determining variability in "intelligence"? Such questions could be raised *ad infinitum* unless an arbitrary limit is set up and adhered to consistently for the purposes of some one particular investigation.

We may conclude from this analysis that the question of the extent of individual differences in different traits cannot be answered unless put in very specific terms. The population must be defined in detail within each investigation and the nature of the trait measured must be made clear, especially by indicating which conditions are to be held

constant and which will be allowed to vary. Obviously all hereditary and environmental conditions which affect a given trait cannot be held constant; otherwise variation would disappear. It should be added that at the present stage in the development of mental testing, owing to the use of incomparable units and arbitrary zero points, the question cannot be answered at all, in any form.

INDIVIDUAL DIFFERENCES IN INFRAHUMAN GROUPS

Individual differences are not to be regarded as characteristically human. Variation is a universal phenomenon throughout the organic scale. "All cats look gray at night," but upon closer inspection each becomes an individual in his own right. Cursory or inadequate observation often creates an impression of similarity or even identity among members of a group, while the differences pass unnoticed. For this reason, only the extreme deviants among animals have attracted attention in the past, all other members of the species having been implicitly relegated to the limbo of "normality."

Several cases of exceptionally "gifted" animals [15] have been described by their trainers or by observers, the remarkable feats of the animals having aroused the wonder and admiration of spectators. Among the most famous examples is Clever Hans, a stallion purchased in 1900 by a Mr. Van Osten of Berlin and subsequently trained by him. The horse was first taught a conventional alphabet in which each letter was represented by a certain combination of taps with the forefoot. Digits were indicated by the appropriate number of taps. By this system, the horse learned to "count" objects presented to him and also to perform all forms of simple arithmetic operations. He could handle fractions, first changing them into decimals. He was able to give the correct answer to such a problem as the following: "I have a number in mind; I subtract 9 and have 3 as a remainder; what is the number?" He seemed to read German readily, and if presented with a series of cards containing written words, he would step up and point with his nose to any words required of him. He answered simple questions put to him orally, tapping out each letter of the answer in his conventional alphabet. He could give the date of any day one might mention, would tell time to the minute, and was able to analyze

[15] For a fuller discussion, see Watson (30, Ch. IX) and Tinklepaugh (28).

a discordant clang, telling his observers which note should be changed.

Most of these feats are not, to be sure, as remarkable as they appear at first glance. Thus, it was found that Clever Hans was unable to respond correctly to a problem if no one present knew the answer. Likewise, when the observers were concealed, the horse failed. The unusual achievements of Clever Hans and of many other performing animals result, not from an understanding of arithmetic or an ability to read, but from an exceptionally keen observation of slight cues given by the observers. The trainer, or other persons present, will make some slight gesture, such as lifting the head a few millimeters, as soon as the animal has tapped the correct number of times. Such cues, it may be added, are usually given unintentionally and unconsciously. They may be too slight to attract the attention of spectators, but an observant animal will learn to respond to them. Although destroying some of the glamour which such feats have had for the public, this explanation does not imply that the task of learning to observe and respond to the proper cues is an easy one which any animal could accomplish.

There remain, furthermore, the cases of animals who have been shown genuinely to respond to a wide variety of verbal commands in the absence of any other cues, or who have learned intricate combinations of movements, or have in many other ways proved their ability to react to very complex situations. Performing dogs, such as "Fellow" who could respond to approximately 400 words and execute the same commands even when worded differently, have been repeatedly exhibited. "Seeing Eye" dogs who lead the blind show a remarkably keen adjustment of their responses to the changing demands of the situation. Chimpanzees have been taught a wide variety of acts, such as skating, riding a bicycle, eating with knife and fork, unlocking doors. The performances of circus animals, and especially "musical" sea lions, are well known. The observation of such animals, even when stripped of popular overstatement, still yields instances of marked individual differences.

Nor is the evidence for individual variation among infrahuman animals confined to the study of unusual cases. Every laboratory investigation employing more than one subject has revealed individual differences.[16] Animal psychologists have not as a rule been concerned

[16] Cf., for example, the discussion of this problem from various angles by Tryon (29).

TABLE 3 *Some Typical Data on Individual Differences in Infrahuman Organisms*

Conditioning Experiments*

Organism	*No. of Cases*	*Conditioning Stimulus*	*Conditioned Stimulus*	*Combinations for Conditioning* Average	Range	SD
Protozoa	82	Tactile	Light	138.5	79–284	24.6
Crustacea	14	Tactile	Light	503	34–1112	
Fish	59	Food	Sounds	12.7	3–35	7.7
Pigeons	13	Shock, Food	Lights, Sounds, Rotation		30–40	
Sheep	11 (estimated)	Shock	Sounds, Tactile		3–17	

Problem Box †

Organism	*No. of Cases*	*No. Learning Step I*	*Trials to Learn Step I* Average	Range	SD	*Range in Steps Learned*
Guinea pigs	30	16	185.50	53–407	176.28	0–1
Albino rats	35	24	221.04	30–453	125.26	0–2
Cats	62	62	46.69 ‡	9–136	25.28	3–7
Monkeys (Rhesus)	17	17	162.47	19–310	94.36	2–22
Monkeys (Cebus)	6	6	137.17	42–327	108.41	5–15

Maze-Learning §

Organism	*Type of Maze*	*No. of Cases*	*No. of Trials to Learn* Average	SD
Albino rats	8 cul de sac elevated skeleton maze	186	32.75	16.59
Albino rats	Equal-unit maze	40	6.40	2.99

* From Razran (18), pp. 308–309.
† From Fjeld (6), p. 528, and Koch (15), pp. 186, 208.
‡ Cf. footnote 17.
§ From Corey (4), p. 256, and Jackson (13), p. 27.

with the measurement of variability, so that the data on this problem are usually mentioned only incidentally and frequently are not given in quantitative form. Whenever such data are reported, however, the range of performance in a randomly selected group is surprisingly large. Wide individual variation has been found in every phase of behavior investigated, such as the amount of general spontaneous activity, the relative strength of drives, emotionality, speed of movement, quickness of learning simple tasks, and behavior in more complex problem-solving situations. Some typical quantitative results on learning behavior have been brought together in Table 3. The average, range, and standard deviation for each set of data have been given whenever available.

The first set of data is taken from experiments on conditioning. Two stimuli, such as a flash of light and an electric shock to the foot, are presented together. After a number of combined repetitions of these stimuli, the withdrawal response becomes conditioned to the light, i.e., the animal will withdraw its foot upon appearance of the light alone, without the presence of the electric shock. It is customary in such an experiment to refer to the original stimulus (in this case, the shock) as the conditioning stimulus, and to the other as the conditioned stimulus. The general nature of the conditioning and conditioned stimuli employed in each experiment has been indicated in Table 3, together with the type and number of animals investigated. It will be noted that the number of combined repetitions of the two stimuli required to establish the conditioned reaction differs widely from individual to individual within each group.

Another set of data is furnished by a series of learning projects conducted at the Columbia University laboratory of comparative psychology. Small samplings of guinea pigs, albino rats, common short-haired cats, and monkeys of two species were tested with the same type of "problem box," in which a series of steps of increasing complexity was presented to the animal. The box consisted essentially of an outer and an inner cage, the latter containing the incentive which the animal obtained at the completion of each successful trial. In the outer cage were three plates to be depressed in a given order by the animal before the door to the incentive compartment was opened. In Table 3, only the number of trials required to learn step I are reproduced, since this was the only step learned by all the groups. The

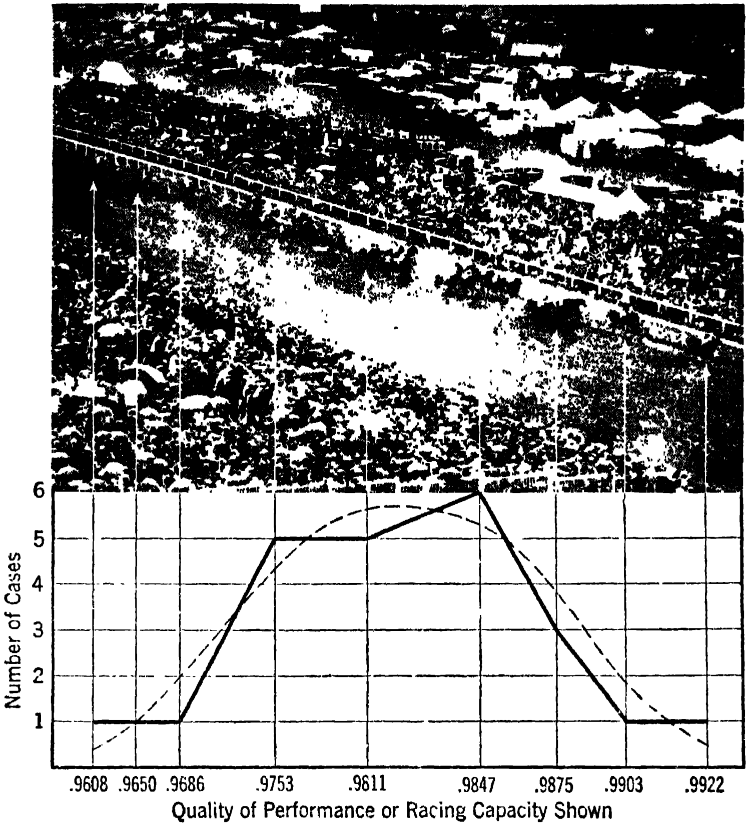


Fig. 32. A Normal Distribution Curve of Racing Capacity, Showing the Field of 24 Horses "Nearing the Line" in the Derby Stakes at Epsom Downs. (After Laughlin, 16, p. 215.)

problem in step I consisted simply in stepping on the first plate to the right as the animal entered the box.[17] The other steps involved stepping on plates 1 and 2; 1, 2, and 3; 1, 2, 3, and back to 2; 1, 2, 3, 2, 1; and so on to other combinations.

Although these problem box studies were conducted mainly to determine the highest number of steps which any animal within a given species could master, the data yield striking evidence of individ-

[17] In the study on cats, the problem set in step I was simpler, the animal being allowed to step on *any one* of the three plates. The data on this group are therefore not directly comparable to those on the other species.

ual differences within each species. Not only the number of trials required to learn each step, but also the number of steps which could be learned, differed from individual to individual. In the group of guinea pigs, some were unable to learn even step 1, while others succeeded; among the rats, some learned two steps, some one, and a few none; among the cats, the range is from 3 to 7 steps; among the rhesus monkeys 2 to 22, and among the cebus 5 to 15. Thus the individual variation was so large that an individual could easily be found in a "higher" species who was unable to learn as much as a given individual in a "lower" species.

In the third section of Table 3 are presented some typical data on maze learning among albino rats. The individual differences are again marked, as is indicated by the standard deviations of the number of trials required to master the correct path in each maze. It is thus apparent that close observation and measurement of animal behavior reveal fully as much individual variability as the studies on human subjects.

An interesting example of the normal distribution curve in a functional trait in animals is to be found in the photograph and accompanying curve reproduced in Figure 32. The photograph shows horses on the race track just before the finish. The relative position of the horses furnishes a vivid demonstration of the normal distribution of racing performance. A few are in the lead, an equally small number lag behind, and the majority are scattered in intermediate positions. The graph is a frequency curve of the "racing capacity" of the same horses, computed by a standardized formula.

CHAPTER

4

Heredity and Environment

WHY DO INDIVIDUALS DIFFER from one another? What are the factors which produce variation? These questions have stimulated prolonged discussion and led to lively controversy. In addition to its fundamental theoretical importance, the problem of the causation of individual differences has far-reaching practical significance in many fields. Any procedure involving the control of human development must be based upon an understanding of the factors which influence such development. All educational methods make some assumption regarding the causes of individual differences. Is the main function of education to produce certain desirable traits, or merely to offer opportunities for the development of the child's "potentialities"? Volumes have been devoted to argumentative and frequently verbose analyses of this question. The empirical accumulation of facts on the causes of individual variation alone can furnish a conclusive answer.

The type of educational activities, vocations, and other pursuits traditionally allotted to men and women rests upon certain beliefs regarding the cause of sex differences in psychological traits. Relationships among racial and national groups, as well as attitudes toward various groups, are based upon theories—either implicitly assumed or overtly stated—regarding the origin of racial and national characteristics. Any caste system implies a hereditary differentiation of people. Although not formally prescribed, such systems still prevail widely, frequently operating in vocational choices and many other situations of everyday life. The interpretation of family resemblances, and even in some cases the development of family groupings themselves, rests upon

specific underlying hypotheses regarding the causal factors in human resemblance and dissimilarity.

THE NATURE OF HEREDITY [1]

The basis of individual differences is to be found in each individual's hereditary background and in the environmental conditions under which he has developed. Let us first consider what, specifically, is meant by "heredity." It need hardly be mentioned, of course, that as herein used the term "heredity" signifies biological heredity. It is only figuratively that we speak of "social heredity," as in such expressions as "the cultural heritage of the twentieth century" or the "inheritance of the family fortune." So-called social inheritance actually falls under the heading of environmental influences.

Basically, an individual's heredity consists of the specific *genes* which he receives from each parent at conception. To call a certain influence, factor, or characteristic hereditary should thus mean that it can ultimately be traced to the presence of a particular gene or combination of genes. The genes are grouped into *chromosomes*, or "colored bodies," so named because they become visible within the cell nucleus when the cell is stained with certain dyes for observation. Chromosomes occur in pairs, the two members of each pair being similar in appearance and function. The number of chromosomes in each cell is, in general, constant within each species, but differs from one species to another. Each human cell, for example, contains 48 chromosomes (24 pairs); in each cell of the mosquito, there are 6 (3 pairs); and in each cell of a certain species of crayfish, there are 200 (100 pairs).

Chromosomes are visible under a microscope, appearing as rod-like, sausage-shaped, or V-shaped bodies (cf. Fig. 33). The genes within each chromosome, however, are so minute as to be invisible, even with a high-power microscope. Through the observation of giant chromosomes which have been discovered within the salivary glands of certain species of flies, it has proved possible to examine the internal structure of chromosomes more fully under the microscope. Al-

[1] To fill out the very brief sketch of the mechanism of heredity which follows the reader is urged to consult any recent text on genetics, such as Sinnott and Dunn (24) or Snyder (25). For a discussion of the concept of heredity, cf. Holt (13), Jennings (14), and Muller, Little, and Snyder (19). A very readable popularized account of heredity is offered by Scheinfeld (22).

though in volume they are from 1000 to 2000 times larger, in other essential characteristics these giant chromosomes are like those found in other body cells. Figure 34 shows a segment of a giant chromosome from the salivary glands of the fruit fly, *Drosophila melanogaster*. Even in such photographs, however, direct observation of the genes themselves is not possible. More recently, the development of the electron microscope, which produces a much higher degree of magnification, has offered new opportunities for investigating the internal structure of chromosomes and may ultimately permit a more direct study of the nature of genes (2).

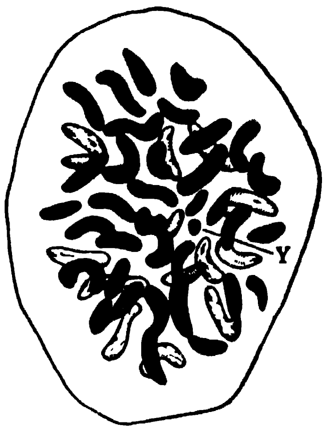


Fig. 33. Human Chromosomes as Seen Under a Microscope. (From Evans and Swezy, 8.)

In the normal process of cell division, or *mitosis*, every chromosome is duplicated by splitting longitudinally along its entire length. Each cell resulting from this division receives an identical set of chromosomes. All cells in the body thus have *identical heredity*. That some develop into eye cells, others into skin, bone, or any of the other varieties of body cells depends upon the influence of the *cellular environment*. Such conditions as gravity, pressure, availability of oxygen and other chemicals, and electrical fields operate differentially upon individual cells, depending upon the position of the cell in relation to other cells.[2] It is believed that the genes, which have been described as "minute packets of chemicals," act as catalysts in these interactions between the cell and its environment.

When the individual has attained sexual maturity, a different type of cell division occurs in the formation of the specialized reproductive cells, the ova of the female and spermatozoa of the male. This process is known as *reduction division*, since the chromosomes in each repro-

[2] Technically, this means that "physiological gradients" of development are established, such as surface-interior, dorso-ventral, or antero-posterior gradients (cf. Child, 6, especially Ch. VIII and IX).

ductive cell are reduced to one-half the original number. Instead of duplicating, as they do in mitosis, the two chromosomes in each pair separate, one going to each daughter cell. It should be noted that in this type of cell division each cell may receive a different combination of chromosomes, since the chromosomes in each pair *assort at random.* Moreover, the chromosomes are not always segregated as units into the different daughter cells, but segments of one chromosome may combine with segments of another ("crossing over"), thus increasing the variety of possible combinations of genes in the individual daughter cells. When the ovum of the mother unites with the spermatozoon of the father in the process of fertilization, the full number of chromosomes is restored and remains through the subsequent mitosis of the developing offspring.

Fig. 34. Giant Chromosome from the Salivary Gland of the Fruit Fly. (From Painter, 21, p. 464.)

The hereditary basis for individual differences is furnished by the almost unlimited variety of possible gene combinations which may occur, especially in such a complex organism as man. It should be noted, first, that even simple human characteristics generally depend upon the combined influence of large numbers of genes. Secondly, the individual germ cells of each parent organism contain different combinations of genes, as a result of the process of reduction division. Thirdly, the cells of two organisms, the mother and the father, combine to produce the new organism, thereby further increasing the variety of possible gene combinations. It should thus be apparent that no two *siblings* (i.e., brothers or sisters) will have identical heredity. The same is true of *fraternal twins,* who, although born at the same time, develop from separate germ cells and are no more alike in heredity than ordinary siblings. Fraternal twins may be of the same or opposite

sex, and may be quite unlike in appearance. *Identical twins*, on the other hand, develop from the division of a single fertilized ovum and therefore have identical sets of genes. Such individuals are complete duplicates as far as heredity is concerned.

The simplest illustration of the mechanism of heredity is furnished by *unit factors*, which depend upon a single pair of genes. An example of such a unit factor is albinism, or the absence of pigmentation in the eyes, hair, and skin. If the individual received a gene for albinism from each of his parents (cc), he will himself be an albino. Individuals with two genes for normal color (CC) will have normal pigmentation. Both of these individuals are described as *homozygous* with respect to albinism. This simply means that the fertilized ovum, or zygote, from which such individuals developed received like genes for albinism or for normal coloring from both parents. If an individual received the gene for albinism from one parent and the gene for normal coloring from the other parent (Cc), he is said to be *heterozygous* in this characteristic. Such an individual will show normal coloring, since normal coloring is *dominant* and albinism is *recessive*. In other words, albinism, being a recessive factor, appears only when the individual has received the recessive gene for albinism from each parent. The heterozygous individual (Cc), although himself normal in coloring, nevertheless carries the recessive gene for albinism, which he may in turn transmit to his offspring.

In the case of other unit factors, the heterozygous individuals may exhibit *blending* rather than dominance. For example, in poultry, black and splashed-white coloring are a corresponding pair of unit factors, but neither is dominant. A cross-breed of these two varieties of poultry will produce individuals of a third color, known as "Blue Andalusians," unlike either of the two parents.

The sex of an individual is itself determined by a pair of chromosomes, known as the sex chromosomes, and designated X and Y. If the child receives an X chromosome from each parent, it will be a female; if one X and one Y chromosome are received, a male will result. From its mother, the child can receive only X chromosomes; while the father can pass on either an X or a Y chromosome. The Y chromosome is relatively small and is believed to contain very few genes. Sex differences in a number of other characteristics may occur because of specific genes carried by the X chromosome. Several such *sex-linked* characteristics have been identified, among the best-known

examples being color-blindness and hemophilia.[3] Both of these conditions depend upon a recessive gene carried in the X chromosome. If a daughter inherits this factor from one parent only, the dominant normal gene in the other X chromosome will prevent the appearance of the defect. Thus a girl will show the defect only if she inherits the defective gene from both parents. In the case of a boy who receives an X chromosome with the defective gene, on the other hand, the defect will invariably appear, since there is no corresponding normal gene in the Y chromosome. Consequently, such characteristics are more common among males than among females.

Certain other factors, such as baldness, are *sex-influenced,* i.e., they behave as dominants in one sex and as recessives in the other. Thus baldness will develop in a male if the gene for baldness was transmitted by either parent. In the female, it will develop only if genes for baldness were received from both parents. Still other factors, known as *sex-limited*, are present in both sexes, but their expression is inhibited in one sex by the presence of the sex hormones. Many of the physical differences between the sexes are probably based upon this type of factor. Destruction or improper functioning of the endocrine sex glands can thus bring about changes in the development of these characteristics.

It should be noted that whenever a characteristic depends upon a single pair of unit factors, the result will be distinctly identifiable types which differ qualitatively from each other. Most traits, however, depend upon *multiple factors,* the number of resulting combinations increasing rapidly as the number of contributing factors increases. With even a relatively small number of contributing factors, the resulting individual differences are quantitative and their distribution may approximate the normal curve. Body height is an illustration of such a multiple-factor characteristic in the human.

In the case of certain multiple-factor characters, the appearance or non-appearance of the character itself depends upon a unit factor. In other words, the operation of the multiple factors is itself dependent upon the presence of a specific gene, which may thus be regarded as a *limiting condition*. The illustration of albinism may again serve in this connection. It is now known that the determination of human

[3] A condition in which the blood fails to clot and the individual may therefore bleed to death even from a slight scratch. This condition attracted especial notice because of its occurrence in certain royal families of Europe.

eye color depends upon the presence of several pairs of genes. Different combinations of such genes produce the almost continuous gradations of observable eye color. If, however, an individual has received the unit factor for albinism from poth parents (cc), he will be an albino regardless of what combination of eye-color genes he may have. The latter are rendered inoperative in the determination of his eye color by the presence of the pair of genes for albinism. Similarly, the spotted coat found in certain breeds of cattle results from a single recessive factor. But the degree of spotting varies along a virtually continuous scale and depends upon a number of modifying multiple factors. This type of relationship is especially relevant to the possible role of heredity in the development of some psychological characteristics. We shall, in fact, have occasion to refer to it again in our discussion of certain types of feeblemindedness (cf. Ch. 16).

Finally, mention should be made of the concept of *"genic balance."* For purposes of analysis, the biologist must necessarily study the influence of particular genes upon the development of each characteristic. We must remember, however, that every characteristic actually results from the interaction of *all* the genes which the individual has inherited. Snyder (25, p. 232) summarizes the contemporary viewpoint of geneticists on this point as follows:

> A gene always exerts its effect in the presence of other genes; hence has arisen the idea of genic balance, by which is meant that any character is the result of the entire gene complex acting in a given environment. Variations in a character may be produced by variations in a single gene, but always in the presence of the rest of the genes.

THE NATURE OF ENVIRONMENT

The concept of environment also requires some clarification. The popular definition of environment is a geographical or residential one. A child is said to have a "poor environment," for example, because he lives in the slums. Or his "environment" is characterized as a French village, an American small town, or a Welsh mining community. Psychologically, such descriptions of environment are highly inadequate. It cannot be concluded, for example, that an 8-year-old boy and his 5-year-old brother standing in the same room at the same time have identical psychological environments even at that moment. The very fact that the current environment of the former includes the presence

of a younger sibling and that of the latter the presence of an older sibling constitutes a significant psychological difference. Moreover, the differing backgrounds of past experience of the two siblings will in turn cause a difference in what each gets out of the present situation. One point is obvious from this illustration: the fact that two children have been brought up in the same home is no indication that they have had identical psychological environments.

Psychologically, environment consists of the sum total of the stimulation which the individual receives from conception until death. This is an active concept of environment, i.e., the physical presence of objects does not in itself constitute environment unless the objects serve as *stimuli* for the individual (cf. 15, 29). This definition is also more inclusive than the popular one, covering all forms of stimulation and extending over the entire life cycle.

The importance of the *prenatal environment* in determining the individual's development has been fully demonstrated. Variations in diet and nutrition, glandular secretions, and other physical conditions of the mother, for example, may exert a profound and lasting influence upon the development of the embryo. That the structural development of the organism is definitely influenced by early environmental factors is clearly indicated by a number of experimentally induced alterations in lower animals.

A curious transformation can be environmentally produced in the axolotl, a large salamander (cf. 14, pp. 117, 124–125). Normally, this animal has prominent external gills, a large tail adapted for swimming, and other characteristics suited for aquatic life. If the young axolotl is fed on thyroid, it loses its gills and its body becomes generally altered so that it is no longer adapted to swimming. The animal then becomes a land salamander, known as Amblystoma, and returns to the water only to lay its eggs.

In the fruit fly, a defective gene causes the animal to produce "reduplicated legs," i.e., certain joints of the legs, or entire legs, are doubled. Although the inheritance of this defective gene has been definitely traced, this characteristic will not appear under certain environmental conditions (12). When animals known to have the defective gene are kept at a sufficiently warm temperature, the additional leg or joint will not develop. Successive generations bred under these conditions will have a normal appearance. If, however, any of their offspring are

allowed to develop in colder temperatures, the defect will reappear. This furnishes a definite illustration of the fact that even a clearly demonstrable "inherited defect" is actually only a tendency to develop in a given way under certain environmental conditions.

Experimentally produced "monsters" represent conspicuous examples of the influence of prenatal environment (26, Ch. VI and VII). In experiments on fish eggs, "siamese-twin" fish have been produced by artificially inhibiting or slowing down the rate of development at an early age through low temperature, insufficient oxygen, or ultraviolet rays. In some cases, one twin is much smaller than the other and is deformed, the larger twin being a perfectly normal fish. Two-headed monsters have been produced among tadpoles and several species of fish by the application of various chemical or mechanical stimuli.

Fundamental variations in the number and position of the eyes of minnows have likewise been artificially induced. If the eggs of the minnow are allowed to develop in sea water to which has been added an excess of magnesium chloride, peculiar eye conditions will appear in a large majority of the embryos. Instead of the usual two eyes, many will develop a centrally placed "cyclopean" eye, so named after the one-eyed Cyclops of mythology. Others may show a single lateral eye, placed to the right or left of the head. Or the two eyes may be abnormally close together. Some of these artificially produced monsters are shown in Figure 35.

Other physical or chemical agents may be employed to produce the same anomalies of development. The primary determining factor in the development of a particular abnormality seems to be the stage at which the agent is introduced, rather than the nature of the specific agent employed. The essential effect is a change in the rate of development, which alters the balance of growth among the different parts of the organism. In commenting upon these experiments, Stockard writes (26, pp. 109–110):

In other words, the genetic composition of these fishes causes them to develop two eyes in normal sea-water, but the same genetic composition gives rise to a single cyclopean eye when an excess of magnesium chloride is added to the sea-water. If sea-water normally had the composition which causes fish to develop with the cyclopean eye, and an experimenter should develop the eggs of fish in a solution of the same composition as

our ordinary sea-water, he would find them giving rise to fish with two lateral eyes instead of the median one, and these two-eyed specimens would appear to this imaginary investigator as monsters.

Thus we cannot even speak of certain structural characteristics as being "normal" for a given species and fixed by hereditary constitution. If the environment in which the organisms develop were to undergo a change of a more or less permanent nature, a different set of characteristics would come to be considered normal. Similarities of develop-

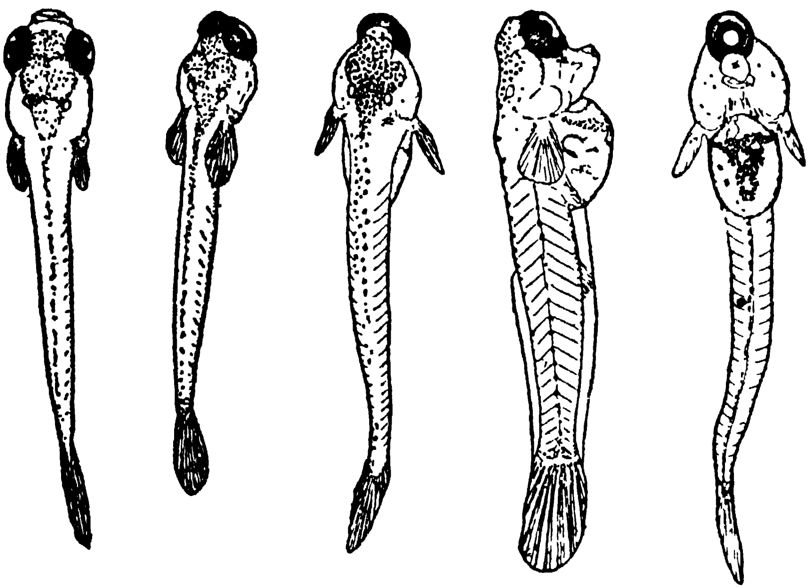


Fig. 35. One-Eyed "Cyclopean" Minnows Resulting from Environmental Conditions. (From Stockard, 26, p. 109.)

ment are attributable to common exposure to an essentially similar environment as much as to the possession of common genes.

Observations on various species have also demonstrated considerable *behavior development* during prenatal life, as well as the influence of specific conditions of the prenatal environment upon such development (cf. 5). The "zero-point" of behavior falls well before birth, the "behavior age," or "mental age" at birth varying widely from species to species (5). Stages of motor development have been clearly established in the embryos of many animals. Sensitivity to

various types of stimuli has also been noted early in prenatal life. Hence the fact that various functions may have been exercised before birth cannot be ignored in the study of subsequent behavior development. The possibility of conditioning to changes in temperature, pressure, and other stimuli in the prenatal environment must likewise be taken into account. Investigation of prenatal learning opens an interesting field of research into the origins of behavior.

Finally, it should be noted that, with increasing precision of definition, the concept of environment has gradually broadened, and that it has also become less sharply distinguishable from the concept of heredity. The popular identification of environment with "external" and heredity with "internal" influences has had to be discarded in the light of increasing knowledge of the operation of heredity and environment. In the preceding section, reference has already been made to *inter-cellular environment,* i.e., the environment consisting of surrounding bodily cells, in which each individual cell develops. The important role of this cellular environment in the establishment of gradients and in other developmental processes is now recognized.

Carrying the analysis still further, we should also consider the *intra-cellular environment.* It is obvious that the genes exert their influence in an environment consisting of the cytoplasm of the cell. The role of the intra-cellular environment is especially important after some differentiation has occurred in the process of cell division. Cells which contain identical genes but different cytoplasmic structure will differ in their ultimate development. The original differentiation occurs under the influence of the genes, but once it has taken place, it in turn affects the further action of the genes.[4] It should be added that each gene must also be regarded as operating in an environment of other genes within any one cell. This mutual interdependence of genes is what is meant by the concept of genic balance, discussed in an earlier section.

From a slightly different angle, mention may be made of the fact that genes themselves, the essential element in any definition of heredity, are not completely immune to environmental influences. Experiments with various types of *radiation,* including X-rays, radium rays,

[4] Geneticists have proposed that the genes may operate as enzymes or catalysts, inducing chemical changes in the cyptoplasm without themselves becoming altered The enzymatic action of a particular gene may produce different results (or no result at all), depending upon the specific chemicals in the cytoplasm of a particular cell. This theory does not preclude the possibility that genes may also exert their influence in other ways.

ultraviolet light, and heat rays, have demonstrated the susceptibility of genes to such influences. Similar effects have more recently been obtained with certain chemicals (7). Since the genes themselves are affected, the changes produced by these agents are not only manifested in the immediate offspring, but are transmissible to future generations. In such a case, a hereditary variant (mutation) results from the operation of an environmental factor. These experiments serve to demonstrate further the fineness of the line which separates the operation of heredity and environment.

THE HEREDITY-ENVIRONMENT RELATIONSHIP [5]

The early concept of "instinct," still prevalent in much popular thinking, implied the existence of behavior which is wholly hereditary. The classification of behavior into "instincts" and "habits," corresponding to "native behavior" and "acquired behavior," respectively, assumed the *exclusive operation* of either heredity or environment within a given activity. Such a theory, implying the hereditary transmission of behavior functions *in toto*, has been quite generally superseded in contemporary psychology. It is now recognized that every trait of the individual and every reaction which he manifests depend both upon his heredity and upon his environment. Although commonly admitted to be untenable, the belief that psychological characteristics can be separated into those which are inherited and those which are acquired is implied in various loosely expressed generalizations about the inheritance of behavior characteristics. Discussions regarding the "inheritance" of intelligence, special talents, or insanity, for example, frequently leave the impression that the inheritance of the behavior itself was meant. Nor are more recent and more sophisticated psychological writings wholly free of such implications. Upon careful consideration, however, it is apparent that hereditary and environmental factors cannot be so glibly separated, nor can behavior be naïvely divided into that which is inherited and that which is acquired.

A second possible way in which the heredity-environment relationship may be conceived is in terms of *additive contribution*. According to this view, both heredity and environment contribute to all behavior development, and the resulting behavior characteristics can be ana-

[5] Much of the present section is based upon a recent article by Anastasi and Foley (1)

lyzed into the *sum* of hereditary and environmental influences. That heredity and environment contribute jointly to the development of behavior is undoubtedly the most widely held view, but the additive assumption regarding their operation is rarely expressed as such. Just this assumption, however, underlies all attempts which have been made to determine the proportional contribution of heredity and environment to the development of particular behavior characteristics.[6] A statement that "heredity contributes 75% and environment 25% to the development of intelligence," for example, would illustrate this additive approach. It might be noted that the same investigators who have offered such estimates of proportional contribution have occasionally argued against the additive view of heredity and environment, apparently unaware of the inconsistency in this procedure (cf., e.g., Burks, 3, 4).

The most widely expressed view of the heredity-environment relationship is that of *interaction*. This means primarily that the effects of hereditary and environmental factors are not cumulative or additive, but rather that the nature and extent of the influence of each type of factor depend upon the contribution of the other. In other words, any one environmental factor will exert a *different influence* depending upon the specific hereditary material upon which it operates. Similarly, any hereditary factor will operate differently under different environmental conditions. It is apparent that any estimate of *the* proportional contribution of a hereditary or environmental factor is inconsistent with this viewpoint, since the proportion would vary as either hereditary or environmental factors varied. To the question, "What is the relative contribution of heredity and environment to individual differences in, let us say, IQ?" there would thus be an infinite number of possible answers.

As an illustration of this point we may consider a non-psychological characteristic whose heredity is known. The number of facets in the eyes of the fruit fly, *Drosophila*, have been found to vary widely in a number of types which differ in their gene constitution. The temperature at which the larvae are kept also determines the actual number of eye-facets which develop. The interaction of these two factors, hereditary and environmental, is illustrated in Figure 36. This graph

[6] An extensive analysis of the implications of the concept of "proportional contribution," as applied to the heredity-environment problem, is to be found in Loevinger (18). Cf. also Schwesinger (23).

shows the effect of temperature upon the number of eye-facets in two types of individuals differing in genetic constitution, which for convenience have been designated "genetic type A" and "genetic type B" on the graph. It will be noted that the *form of the curve* differs for the two genetic types. The difference in number of eye-facets between the

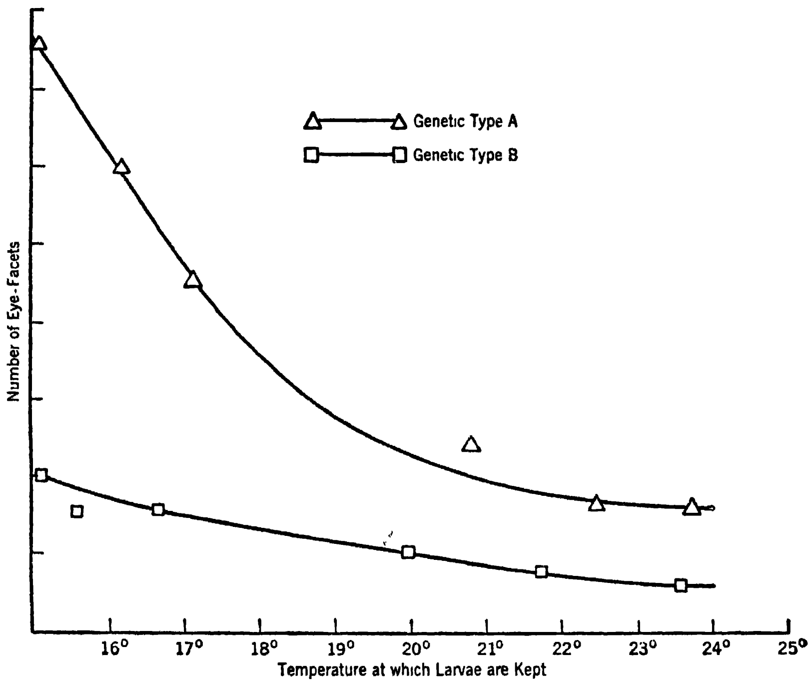


Fig. 36. An Illustration of the Interaction of Hereditary and Environmental Factors: Number of Eye-Facets in *Drosophila* as a Function of Genetic Constitution and of Temperature. (From Hogben, 11, p. 96.)

two genetic types was much greater at 16° than at 25°. Conversely, the effect of temperature was greater on one genetic type than on the other. Thus, a "different difference" resulted from environmental changes when operating on individuals of different heredity; and a "different difference" resulted from hereditary variations when operating in different environments. The "ratio" of hereditary and environmental contributions would thus vary as either factor varied.[7]

[7] For other illustrations, cf. Haldane (10) and Hogben (11).

The operation of a similar type of heredity-environment relationship can readily be recognized in the development of many familiar human characteristics. If we ask, for example, to what extent body weight depends upon such environmental factors as diet and exercise and to what extent it depends upon hereditary factors, no single answer can be given for all individuals or all environmental conditions. Because of differences in hereditary factors, the body weight of certain individuals is more susceptible to differences in diet, exercise, etc., than that of other individuals. In the former type of person, the contribution of heredity is smaller. Thus, the proportional contribution of heredity and environment to body weight may itself be determined by hereditary factors, and may vary from person to person. For example, for some individuals diet and other environmental factors might contribute 10% to the determination of body weight, for others 80%. The proportional contribution of heredity and environment may likewise be altered by variations on the environmental side, such as the absolute amount of food intake. Thus when the total amount of food intake is low, as in a near-starvation diet, body weight undoubtedly depends to a much greater extent upon differences in the amount of food. When the total intake of food is large, individual differences in body weight are probably much less dependent upon diet.

Finally, we may consider a hypothetical illustration involving intelligence test scores. Suppose we find a 10-point difference in IQ between two identical twins reared in separate foster homes (A and B), and a 30-point difference in IQ between two unrelated children, one reared in foster home A and the other in foster home B. Can we argue that the 10-point difference between the identical twins measures the "differentiating effect" of these two home environments, and that the 30-point difference between the unrelated children can therefore be analyzed into 10 points attributable to environment and 20 points attributable to heredity? Could we conclude that, in so far as these cases show, heredity is twice as important as environment in the production of individual differences in IQ? If we follow the concept of interaction, the answer to both questions is "No." Actually, a very slight hereditary difference between the two unrelated children may have greatly augmented the difference between the effective environments of the two foster homes. The difference in environmental stimulation between the two homes would thus have been much greater for the unrelated children than for the identical twins. No simple sub-

traction of the end-products could disentangle the relative contribution of the factors whose initial interaction led to the obtained difference in IQ.

All the examples of interaction which have been discussed—the eye-facets of the fruit fly, as well as the hypothetical examples of body weight and IQ—illustrate the *interdependence* of heredity and environment, which is fundamental to the concept of interaction. To summarize, interdependence means that the contribution of any given environmental factor to a particular trait depends upon the individual's specific hereditary background; and conversely, the contribution of any given hereditary factor depends upon the specific environmental conditions within which it operates. Another implication of the concept of interaction is that the heredity-environment relationship can be more accurately likened to the arithmetic operation of *multiplication* [8] than to that of addition. The individual's characteristics may be conceived as the product, rather than the sum, of the hereditary and environmental factors. Under these conditions, a slight difference in environment, in combination with a slight difference in heredity, may ultimately lead to a very large difference in the resulting characteristic. We must envisage such a "multiplication" of influences as occurring successively in the individual's development, each new "product" being itself the basis for further multiplication in an ever-widening radius. Thus a slight initial difference between two individuals may launch them on two widely diverging paths of development.

Still another implication of the concept of interaction should be recognized. Any estimate of the relative contribution of hereditary and environmental factors to individual differences obviously depends upon the *range* or extent of both hereditary and environmental differences within the population under consideration. For example, susceptibility to diphtheria has been shown to depend upon a recessive hereditary factor, and immunity upon a corresponding dominant factor (25, pp. 370–371). This disease will not be contracted, however, without infection by the diphtheria bacillus. If, now, we consider a population all of whom have inherited susceptibility, then individual differences in the development of the disease could be attributed entirely to the

[8] To speak of hereditary and environmental factors as being multiplied is obviously an oversimplification, although helpful in visualizing the relationships involved. The actual mathematical function by which hereditary and environmental contributions combine is unknown and may well differ from one specific characteristic to another.

environmental differences, i.e., exposure to infection. On the other hand, in a population in which all are equally exposed to the bacillus, any individual differences would be attributable to differences in heredity, i.e., whether the dominant gene for immunity was present. To the question, "What proportion of the variance in the development of diphtheria is attributable to heredity?" opposite answers would be reached in these two populations. Similarly, a wide variety of intermediate answers could be reached in other populations, depending upon the relative frequency of exposure and the relative frequency of individuals with the dominant gene for diphtheria immunity in each population.[9]

Throughout this discussion, the terms "heredity" and "environment" have frequently been used without qualifications for the sake of brevity. It should not be concluded, of course, that they refer to single entities or forces. Both heredity and environment are general names for *complex manifolds of many specific influences*. In the development of the individual, interaction occurs *within* as well as *between* the specific factors in each of the two categories. To speak of all the thousands of genes, each with its specific chemical and other properties, as though they represented a single force, operating as a unit to stimulate development in a particular direction, is highly misleading. It is even more clearly apparent that "environment" is not an entity which can be contrasted or juxtaposed with "heredity." Cellular environment, radiation effects upon genes, birth injuries, educational history, and socio-economic level can scarcely be treated as a single influence!

POPULAR MISCONCEPTIONS REGARDING HEREDITY AND ENVIRONMENT

A number of misconceptions regarding the operation of heredity and environment are still prevalent in popular thought. Before proceeding farther, we shall examine briefly some of the most common of these

[9] The three implications of the concept of interaction discussed above are independent of each other although all three are generally implied in current discussions of the heredity-environment relationship. It would be logically possible, for example, for the operation of heredity and environment to be additive, while the first and third conditions discussed still held. In this case, estimates of proportional contribution would still be meaningless. Or the heredity-environment relationship might be one of multiplication, without interdependence, i.e., the weight of the hereditary and environmental factors would vary independently of each other.

erroneous beliefs, in order to clear the way for further analysis of the heredity-environment problem.

Hereditary *versus* Inborn. One of the most common sources of confusion in discussions of heredity and environment is that between "hereditary" and "inborn." The popular belief that whatever is present at birth is necessarily inherited is bolstered by the lack of precision in terminology. The dictionary definitions of such terms as "hereditary," "inborn," "innate," "congenital," and "native" are difficult to differentiate. Certainly the terms are often used interchangeably, in the scientific as well as the popular literature. The scientist usually employs all or most of them as synonymous with "hereditary." [10] The layman, on the other hand, frequently interprets all these terms with reference to birth, a reference which is obviously present in the root of such words as "inborn," "native," and "innate."

It is, of course, just as incorrect to regard the influence of heredity in the development of any trait as ceasing at birth as it is to date the onset of environmental influences from birth. Hereditary factors may affect the development of the individual long after birth and, in fact, throughout the life span. Inherited susceptibility to various diseases, for example, may not be manifested until well past middle age. Even the age at which a person dies may be determined partly by hereditary factors, as suggested by the observation that longevity tends to run in families. Hereditary influences may thus become manifest for the first time at any age. That environmental influences begin to operate long before birth has already been demonstrated in the discussion of prenatal environment. The influences of heredity and environment are co-extensive in time. Birth is not to be regarded as either a beginning or an end in the operation of these factors, but as one event in a developmental continuum which for the individual begins at conception and ends at death.

Resemblance to Parents. Another popular fallacy is the belief that heredity means parental resemblance, and vice versa. Both sides of this proposition can be shown to be false. That heredity need not result in the resemblance of offspring to immediate forbears is apparent from a consideration of the mechanism of heredity. The genes are continuous from generation to generation. They are not "produced" by

[10] In some writings, "congenital" is used to signify presence of a characteristic at birth, as distinguished from "hereditary." There seems to be no linguistic justification for singling out this particular term for reference to birth.

the individual parents, but are simply transmitted by them to their offspring. Thus the individual inherits not only from his parents but also from all his direct ancestors. A characteristic which has remained latent for many generations may become manifest because of a particular combination of genes, e.g., two recessives. The result will be an individual unlike his parents or immediate forbears in some one respect. Instances of this sort are common in family histories. One of the most familiar illustrations is that of two brown-eyed parents having a blue-eyed child, through the combination of two recessive "blue-eye" genes in the offspring. In such cases, heredity actually serves to make the child unlike his parents.

The converse proposition, that parent-child resemblance is necessarily indicative of heredity, is equally untenable. Such resemblances may have developed through the many environmental contacts and similarities of parent and child, both prenatally (in relation to the mother) and postnatally. Not only are parents and children exposed to more nearly similar environments than are unrelated individuals, but they constitute in part each other's environment. Thus mutual influence as well as common stimulation may serve to produce resemblances. For these reasons no parent-child likeness can be attributed to hereditary factors without further analysis of its development.

Inheritance of Acquired Characteristics. The Lamarckian hypothesis of the inheritance of acquired characteristics has found no support either in the experimental findings of genetics or in the data of embryology regarding the mechanism of heredity. Yet the popular belief persists that parents may transmit to their offspring physical as well as psychological characteristics which the parents have developed through training or experience. For example, the opinion may be expressed that if the parents attend college, their children will as a result "inherit" superior mental ability; or that if the parents engage in athletic activities, their children will have stronger muscles. Statements are also made to the effect that the parents' acquired fears, interests, prejudices, ethical or aesthetic standards, mechanical skills, and the like, may be inherited by the offspring.

The truth of the matter is, of course, that only conditions which act directly upon the gametes, or germ cells, are transmissible to the offspring. It is theoretically possible, to be sure, that certain activities of the parents may bring about the operation of effective physical agents upon the genes. Exposure to radiation (as from atomic bombs!)

would be an example. The action of various types of radiation in producing gene modifications, or mutations, was mentioned in an earlier section. Genes are, however, extremely stable, and the agents which affect them very few. Certain other agents, such as alcohol, may injure the cytoplasm of the germ cells, thus affecting the development of the immediate offspring, but producing no inheritable change which might be transmitted to subsequent generations. Such direct physical effects on genes or cytoplasm are, however, a far cry from the "transmission" of an interest in the classics or a taste for non-objective paintings!

"Maternal Impressions." An even more naïve notion pertains to the influence of the mother's experiences during pregnancy upon the characteristics of the child. Under this heading would be included certain popular explanations of "birthmarks" or the superstition that a man may have bushy eyebrows because his mother was frightened during pregnancy by a shaggy-haired Airedale! Another favorite illustration is that of the mother who attends lectures, concerts, and recitals during pregnancy in order that her child may acquire a desire for "culture." All such beliefs are now in the category of superstitions and old wives' tales.

The only prenatal influences which the mother's activities can exert upon the developing offspring are indirect, biochemical effects. Thus certain toxic materials, germs, or any other agents carried by the blood stream can be transmitted by the mother to the embryo. Similarly, the mother's general level of metabolism, her nutrition, and her endocrine balance may exert considerable influence upon the development of the embryo. It follows that excessive emotional excitement during pregnancy, for example, may have an indirect effect upon the developing child, as a result of chemical changes in the maternal blood stream. But there is certainly no basis for expecting specific fears or other experiences of the mother to have a specific physical or psychological effect upon the embryo.

"STRUCTURAL" AND "FUNCTIONAL" CHARACTERISTICS

Up to this point, we have been discussing the operation of hereditary and environmental factors in general, without special reference to psychology. We may now turn to a consideration of the applications of these concepts to psychological phenomena. The proper domain of

psychology is the behavior of individuals. Structural characteristics are important in this connection in so far as they impose certain limitations upon the development of behavior. A cat cannot learn to fly because it has no wings. If a child has a defective thyroid, his movements will be slow and sluggish, and his general behavior dull and stupid. For the development of certain types of behavior, vocal organs, hands, and a human nervous system are essential prerequisites. The nature and development of bodily structures obviously play a part in determining the characteristics of behavior.

The presence of certain structural characteristics should, however, be regarded as a necessary but not a sufficient condition for the development of any specific type of behavior. In other words, the presence of all the structural prerequisities does not in itself insure that the given behavior will appear. It also follows that the absence of a given type of behavior does not necessarily imply a structural deficiency, nor do behavior variations necessarily imply corresponding structural variations. Except for individuals with gross pathological defects, the structural equipment of most persons is such as to permit an almost unlimited variety of behavior development.

Much confusion and controversy in discussions of heredity and environment in psychology arise from a failure to distinguish between behavior characteristics and structural characteristics. Statements regarding the "inheritance" of feeblemindedness, musical talent, mathematical aptitude, or criminal tendencies are at best highly misleading.[11] Certainly, no one expects disembodied functions as such to be mysteriously transmitted through the genes. The genes are obviously specific chemical substances which, through many successive interactions with other substances in the environment, eventually bring about the development of the structures making up the individual. No "potentialities," "tendencies," "influences," "determiners," or other mystical entities can be discovered in the genes.

What, then, can be said regarding the role of heredity in *behavior*? Above all, it is clear that hereditary factors cannot affect behavior directly, but only indirectly through the structural equipment of the individual. The immediate question thus resolves itself into a consideration of the role of structural characteristics in behavior development.[12] In what way are given behavior characteristics related to

[11] In many instances, of course, they are completely unfounded. But at this stage we are not considering the factual material.

[12] For a fuller elaboration of this point, cf. Anastasi and Foley (1).

structural conditions, such as glandular defects, pathological brain conditions, chemical composition of the blood, and the like, and in what way are they related to *functional conditions,* i.e., the individual's previous reactional biography?

When a specific structural condition is found to be associated with a given behavior characteristic, then the question of heredity and environment can be raised. If, for instance, a particular behavior deficiency is shown to be regularly associated with a certain brain condition, this condition may in turn be traceable to the presence or absence of a specific gene or combination of genes. On the other hand, the brain condition may result from physical or chemical characteristics of the prenatal environment, from birth injuries, or from other environmental factors. Lack of one specific gene may prevent normal brain development and thereby result in a form of feeblemindedness. In such a case, this particular type of feeblemindedness would appear as a simple Mendelian unit in genetic studies of family pedigrees. Findings such as these would not, however, justify the assertion that "feeblemindedness" is a simple Mendelian recessive, as was proposed in some of the earlier psychological writings (cf., e.g., 9). In the first place, such a finding does not imply that only one gene is required for normal mental development. Undoubtedly many genes contribute to the structural development necessary for "intelligence." The absence of one or a few specific genes may, nevertheless, prevent the effect of the others from being manifested. Hence a particular defect in a structural characteristic may be transmitted as a Mendelian unit, although the characteristic itself depends upon the combined effect of a large number of genes. In the second place, the presence of all the required genes would not insure normal intelligence. Intellectual development—as all psychological development—depends upon the individual's reactional biography, viz., upon what he does with his structural equipment.

THE CONCEPT OF "UNLEARNED BEHAVIOR"

One of the major sources of confusion and controversy in psychologists' discussions of heredity and environment centers around the con-

cept of "unlearned behavior." Among the criteria for the identification of such unlearned behavior which have been proposed from time to time may be mentioned: universality within a species, uniformity among different members of the species, sudden appearance without subsequent change, uniformity of developmental sequences in those cases in which change does occur, and "adaptiveness" or effectiveness far in excess of that which could reasonably be expected from the animal's own learning. Objections have been raised to each of these criteria (cf., e.g., 17; 20, Ch. III), the principal criticism being that behavior which meets any or all of these specifications can and does at times develop through learning.

The only completely dependable criterion of unlearned behavior is the demonstrated absence of the opportunity to learn. If this criterion is applied, instances of unlearned behavior can still be found in various species, the clearest illustrations being furnished by the behavior of certain insects. In such illustrations, highly uniform and complex series of activities are performed despite the fact that the animal has had no previous contact with other members of the species or with the objects toward which the behavior is manifested. In many such species the parents die or abandon the eggs long before they are hatched. Thus the offspring have no opportunity to learn by observing the parent's behavior, nor does the parent have any opportunity to observe the effect of its preparatory activities upon the offspring.

A favorite illustration of such unlearned behavior is the frequently cited pollinating behavior of the yucca moth. As soon as this insect emerges from its chrysalis, it travels to a yucca flower, from which it obtains pollen. It then finds another yucca flower, where it deposits its eggs as well as the newly gathered pollen, following a highly stereotyped sequence of reactions. The fertilized ovules of the flower, which result from this pollination, provide food for the yucca larvae when they emerge from the eggs four or five days later. In commenting upon the unlearned nature of this pollinating behavior, Stone (27) has written:

> The adult does not partake of the pollen which it gathers and probably obtains no nourishment at all from the plant while performing this round of complicated activities. . . . The adult insect does not learn this complicated series of acts through imitation of its parents, long since dead, or from contemporaries either, for its visual receptors do not provide the

kind of vision necessary to the human concept of visual guidance. Action systems of the larvae are totally unlike those of adults, and the activities are even performed with different appendages. The body of this same larva that descended the silken thread to bury itself in the ground is dedifferentiated and resynthesized during the resting state, and a prolonged interval of time, the winter season, intervenes between the last act of the larva and the first of the adult. In view of these facts, no concept of memory or transfer of training supported by experimental evidence can be invoked to account for the behavior of the yucca moth (?7, p. 46).

Obviously, "unlearned behavior" can only mean behavior which is determined wholly by the structural characteristics of the organism, such that the mere presence of the necessary structures at a certain stage of development insures the appearance of the behavior in question. Merely to say that a certain type of behavior is unlearned, however, is no answer to the question of how it develops. Such a statement only reformulates the problem, so that the question still remains to be answered. The answer now calls for knowledge of what structural factors determine such behavior and how they operate. To prove that behavior is unlearned, i.e., *not* learned, is a negative finding, which furnishes no positive information. It does not in itself tell us how the behavior develops. To call such unlearned behavior "instinctive," "innate," or "hereditary" simply obfuscates the problem, because these terms seem to suggest positive explanations or active processes, whereas *in this case they are being used only as synonyms for the negative term "unlearned."*

The same difficulty arises in the common use of the term "*maturation*" in psychological writings. In discussions of the origin of behavior, a distinction is usually made between development through learning and development through maturation. The latter refers to the sudden appearance of certain behavior, regardless of the previous activities of the organism, as soon as the requisite stage of structural development is attained. This term is misleading for several reasons. It suggests a positive process of behavior development, without making it sufficiently clear that it is the structures that are developing. Moreover, certain writers who use the term "maturation" easily slip into the implication that such behavior results from an "unfolding of potentialities" which were present in the genes, and that it is therefore inherited.

Strictly speaking, it is incorrect to regard unlearned behavior as hereditary. In the first place, behavior cannot be inherited as such. It is only structural characteristics which can be directly influenced by the genes. In the second place, the structural conditions which determine such unlearned behavior may themselves result from either hereditary or environmental factors, or varying combinations of the two.

Certain psychologists maintain that structurally determined or "unlearned" behavior falls outside of the scope of psychology. This is the position taken by Kantor (16, Ch. IV), for example, who holds that biological functioning follows directly from the structural properties of the organism and the physical characteristics of the stimulus, whereas psychological functioning depends upon the individual's previous interactions with stimuli.[13] Other psychologists would be reluctant to exclude consideration of such unlearned, structurally determined behavior from the proper domain of psychology; some have, in fact, devoted virtually all their research to its study. Whether one defines psychology so as to include structurally determined behavior, or whether one insists that all such behavior belongs under the heading of biology is in itself only a question of division of labor or personal interest. The essential point is to have a clear and unambiguous understanding of what is meant by unlearned behavior and to avoid ill-defined, mystical implications in its discussion.

"Unlearned behavior" has been traditionally subdivided into such categories as tropism, reflex, and instinct. These distinctions are not sharply drawn. Some writers have, in fact, used one or another of these terms exclusively to designate all unlearned behavior. The most common usage, however, is to designate as *tropistic* any behavior which is primarily an orienting (turning, approach, withdrawal) response of the entire organism toward a stimulus, such a response being essentially "forced" by the physical and chemical properties of the stimulus and of the reacting organism. An example is the turning and bending of plants toward the sun or other source of light. "*Reflex*" generally refers to a specific response of a part of the organism to a particular form of stimulation. The term is usually applied only to organisms which have a synaptic nervous system. The structural basis of the

[13] It should be added that the stimulus itself is defined differently by Kantor in these two situations (cf. 15, vol. I, Ch. II).

reflex is the "reflex arc," consisting of receptor, neurones, and effector. Two examples in man are the patellar reflex, or "knee jerk," and the pupillary reflex, or contraction and expansion of the pupil as intensity of illumination changes.

The term *"instinct"* has been used with more varied meanings, although nearly all its definitions imply a greater complexity of behavior than is represented by either "tropism" or "reflex." Some use the term to refer to a chain or integration of reflexes, as illustrated by the complex stereotyped sequences of unlearned activities observed in certain insects, such as the yucca moth cited above. Others use the term in a vaguer sense to mean a relatively rough framework within which considerable variability of specific behavior may occur. In such definitions, instinct is often related to physiological needs, such as the need for food or water, and to the presence of hormones. It is this latter, less specific use of the term "instinct" that has opened the way for many unbridled leaps into an improbable terrain. It is here, for example, that one finds discussions of gregarious or collecting "instincts" and the like. Not only have the structural properties leading to gregariousness or collecting behavior never been identified or even vaguely guessed, but the nature of this behavior is also such as to make the search for its structural correlates appear futile and meaningless.

It is undoubtedly true that isolated instances of behavior can be found which clearly fit the definitions of tropism, reflex, or instinct. On the other hand, most behavior—human or infrahuman—cannot be classified into any one of these categories. Certain segments or aspects of a complex activity could probably be described as tropistic, reflexive, instinctive, or learned, the activity itself including more than one of these various components. It would seem, moreover, that these terms, as well as the term "maturation," lend themselves too readily to misunderstanding and unwarranted implications. To say that a given activity or a particular component of an activity is unlearned (provided it has been conclusively demonstrated to be unlearned) is certainly a more precise and objective description of the actual observations. To call such an activity structurally determined adds to the observation the only possible source of the occurrence of such behavior. At the same time, the designation "structurally determined" centers attention on the question which logically follows, viz., what structures are involved and how do they bring about such behavior?

METHODS FOR THE STUDY OF HEREDITY AND ENVIRONMENT

We may conclude this preliminary introduction to the heredity-environment problem with a brief overview of the methodology which has been developed for its study.[14] Psychologists have followed many and varied approaches in their efforts to disentangle the factors which underlie behavior development. Some of the methods yield results which are highly ambiguous and difficult to interpret. Few approaches can give conclusive answers by themselves. Each has its own peculiar advantages and limitations. Frequent resort is made to the combination of methods and the mutual corroboration of data in the attempt to remedy the shortcomings of any one technique. A few of the methods which will be outlined below were not specifically designed for the study of heredity and environment. For example, investigations of sex differences and racial differences are conducted primarily because of an interest in the psychological characteristics of the specific groups under consideration. Some of these studies are nevertheless set up in such a way as to contribute toward an analysis of the factors determining behavior development, and they have been included in the present listing for this reason.

It has often been pointed out that *the* crucial psychological experiment on heredity and environment has yet to be done. The chief difficulty confronting the investigator in this field is that of isolating the influence of hereditary and environmental factors. As in all experimental design, the essential prerequisite is the control of conditions in such a way that all variables shall be constant except the one whose influence is being investigated. Since in most investigations on individual and group differences, hereditary and environmental factors have varied simultaneously, the results are incapable of definitive interpretation.

If heredity can be assumed to be constant, as in the case of identical twins, then differences can be attributed unambiguously to environment. Similarly, if environment is held constant, any observed differences must be the result of hereditary influences. In view of our dis-

[14] A survey of promising methodological resources of which little or no use has so far been made is given by Stone (28). Most of the methods discussed by Stone are limited to relatively simple animal forms.

cussion of what constitutes the effective environment, however, it should be apparent that it is extremely difficult to hold environment constant for two individuals, especially for human subjects. A few of the techniques employed clearly make no attempt to separate hereditary and environmental factors, and consequently yield results which are at best descriptive, and not explanatory. On the other hand, some of the approaches make it possible to segregate, at least partially, the relative contribution of heredity and environment to the development of individual differences in specific characteristics.

For convenience, the principal methods employed by psychologists in studying the heredity-environment problem have been grouped into fourteen categories. A discussion of each of these methods, together with illustrative data, will be found in the appropriate chapters in Parts II and III. These chapters have been indicated in parentheses next to each category.

Since the now famous experiments of Mendel, geneticists have made constant use of *selective breeding* (Ch. 5) to investigate the inheritance of structural characters. In recent years the method has been applied to the study of psychological characteristics. Laboratory rats, for example, have been bred for maze-learning proficiency and other behavior characteristics. This method is obviously not feasible with human subjects.

Normative developmental studies (Ch. 5 and 9) are observational studies of the course of behavior development in the growing organism. Regularity of sequence in developmental stages is of special interest in connection with the heredity-environment problem. Observations of this type have been made on both infrahuman and human subjects, and during prenatal as well as postnatal periods. Studies of the "growth curve" of various behavior functions, and of the decline of functions with age may be regarded as extensions of this method to later age levels (Ch. 9). A closely related approach is the study of the *structural correlates of behavior development* (Ch. 5). This method has been used principally with lower animals and during the prenatal stages, although it is also applicable to a limited extent postnatally and with human subjects. By means of this method, the first appearance of certain behavior functions, such as specific types of movement, may be linked with a particular phase of development in the nervous system or other bodily structures, and subsequent behavior development may be traced in conjunction with structural changes.

One of the most direct approaches to the heredity-environment problem is the *experimental variation of environmental conditions.* This approach is illustrated by the artificial prevention of the exercise of a function and the subsequent observation of the effects of such deprivation upon the development of the function (Ch. 6). Sometimes the experimental variation consists in giving additional exercise or training in a particular function, in order to determine the extent to which the normally observed course of development may thereby be altered. Certain investigations take advantage of a sort of "unplanned experiment" of this type afforded by the varied *infant-rearing practices* of different cultures (Ch. 6). For example, the prevention of locomotion or of the exercise of other motor functions, beyond the age when such functions are well developed in other cultures, permits an analysis of the relative dependence of these functions upon structural growth and upon exercise. A similar "unplanned experiment" is furnished by the cases of so-called *feral man* (Ch. 6). This is a term applied to children who have apparently been isolated from human beings at an early age and have either been reared by animals or have shifted for themselves in the absence of any companions.

A much more restricted type of training experiment is represented by investigations of the *effects of practice and coaching on mental test performance* (Ch. 7). Such studies cover the influence of practice upon the extent of individual differences, as well as the effects of coaching or of the repetition of a test upon the general level of performance. In recent years, considerable attention has been attracted by a related type of investigation, concerned with the *effects of schooling upon mental test performance* (Ch. 8). A large number of these studies have been concerned with the possible increase in IQ following nursery school attendance, although a few have dealt with the influence of education at the elementary school and higher scholastic levels.

Family resemblances and differences have long been a favorite method for the study of heredity and environment, although this approach is beset with many difficulties and its results are likely to prove ambiguous. The study of *family history* (Ch. 10), introduced by geneticists, is most fruitful when applied to relatively simple characteristics. Very limited use has been made of it in the study of psychological characteristics; some of the applications of the method in this area are open to serious criticism. *Intrafamilial correlations* (Ch. 10) have been frequently computed with the results of mental tests admin-

istered to parents and children, siblings, and other related individuals. The study of *twins and foster children* (Ch. 11) offers the opportunity for a more direct analysis of the influence of hereditary and environmental factors.

Investigations of the *relationship between individual differences in structural and behavioral characteristics* (Ch. 12) may suggest certain physical correlates of individual differences in behavior, which could in turn be traced to hereditary or environmental factors. This method is to be distinguished from the direct study of the structural correlates of behavior development, cited earlier. The present method is a statistical rather than a developmental one, and has generally been applied to fairly complex functions in the human adult. Typology, or the search for "constitutional types" with structural as well as psychological differentia, may be included under this heading (Ch. 13).

The *comparison of socio-economic groups* (Ch. 23), including occupational levels, urban and rural groups, individuals living in isolated or "culturally backward" communities, and the like, represents another approach to the problem. A related method is the *cross-comparison of cultural and biological groupings* (Ch. 22). Groups which are biologically differentiated, such as the two sexes or different races, have been compared on a wide variety of psychological tests. Of special interest, however, are those investigations in which "cross-comparisons" can be made between such biological (i.e., hereditary) groupings and the cultural (i.e., environmental) groupings which cut across them.

At this stage in our treatment of the problem, we have considered the concepts of heredity and environment, the complex relationships between hereditary and environmental factors, popular misconceptions in this area, and the importance of distinguishing between structural and functional characteristics. Various implications of the concept of unlearned behavior have also been examined, followed by an introductory listing of the methods used by psychologists in studying the heredity-environment problem. From this preliminary survey, it should be apparent that the problem is by no means a simple one. Alluring generalizations can only mislead in a topic which is intrinsically complex. If this discussion has given the reader some conception of the complexity of the heredity-environment relationship, it has served its purpose well. Moreover, if the reader has come to recognize the importance of careful use of terms, to distinguish between superstition

and established fact, and to follow deductions logically and objectively in the heredity-environment area, he will have made significant strides in his thinking. An honest, forthright recognition of the complexity and inherent difficulties of the problem, as well as the limitations of our present knowledge in this field, is to be preferred to a list of glossy oversimplifications.

PART II ANALYSIS OF INDIVIDUAL DIFFERENCES

CHAPTER 5

Biological Factors in Simple Behavior Development

PART I SERVED AS AN INTRODUCTION to the basic concepts and methodology of differential psychology. We may now consider some of the principal *findings* regarding the nature and sources of individual differences. Throughout the chapters which follow, the persistent question of heredity-and-environment will be repeatedly encountered. The various methods for the study of this question, outlined in the preceding chapter, will be treated in the remainder of the book in connection with the topics to which they are most relevant. The organization of the following chapters is based primarily upon the traditional areas of investigation within differential psychology rather than upon a logical analysis of the problems. Such an organization was obviously necessitated by the available results.

In the present chapter, typical findings pertaining to the development of simple behavior functions will be considered. The studies to be included are classifiable under the first three of the methods listed in Chapter 4. Although these methods have many points of difference, and some could more logically be combined with methods to be treated in later chapters, they have been brought together in this chapter because of certain common features which make their joint consideration convenient. In the first place, these approaches have been concerned with relatively simple behavior to the almost complete exclusion of the more complex linguistic and other symbolic activities. These methods thus make virtually no use of psychological tests, which have played so large a part in many of the other approaches. Since special methodological problems are presented by psychological tests,

it is more expedient to consider separately those investigations which do and those which do not employ tests.[1]

In the second place, the studies to be considered in the present chapter have dealt largely with relatively simple organisms, viz., those at lower phylogenetic or ontogenetic levels. Many of the observations have been made on infrahuman animals. In those investigations employing human subjects, the factors studied operated at an early age, prenatally or during infancy or early childhood. A third feature which characterizes the present group of studies is their emphasis upon the biological or structural conditions underlying behavioral differences. Thus selective breeding, the charting of progressive age changes in behavior, and the observation of structural changes which parallel behavior changes may all be regarded as ways of determining the effect of structural conditions upon behavior development. To be sure, psychological differences in the stimulating conditions are also present when comparing, for example, organisms at different age levels, but the emphasis in the present approaches has been put upon the biological conditions.

SELECTIVE BREEDING FOR BEHAVIOR CHARACTERISTICS

The experimental breeding of animals selected on the basis of behavior characteristics is a recent application by psychologists of the technique of selective breeding long in use by geneticists. This is the basic method of genetics for the study of the inheritance of any characteristic. Through several refinements of this method (as in the analysis of crossovers and "linkage groups"), geneticists have succeeded in analyzing the hereditary basis of many structural characteristics and even in constructing theoretical "gene maps" for certain species. The present applications of selective breeding to behavior phenomena, however, are far from reaching such refinements. All that these studies have achieved to date is to effect through successive generations of selective breeding, the development of two strains which differ significantly in a given behavior characteristic. Following the establishment of the contrasted strains, a beginning has been made in the search for possible structural bases for such behavior differences between strains.

[1] The method of selective breeding, for example, could logically be combined with studies on familial resemblances and differences, to be treated in Chapter 10, but the latter approach makes extensive use of psychological tests.

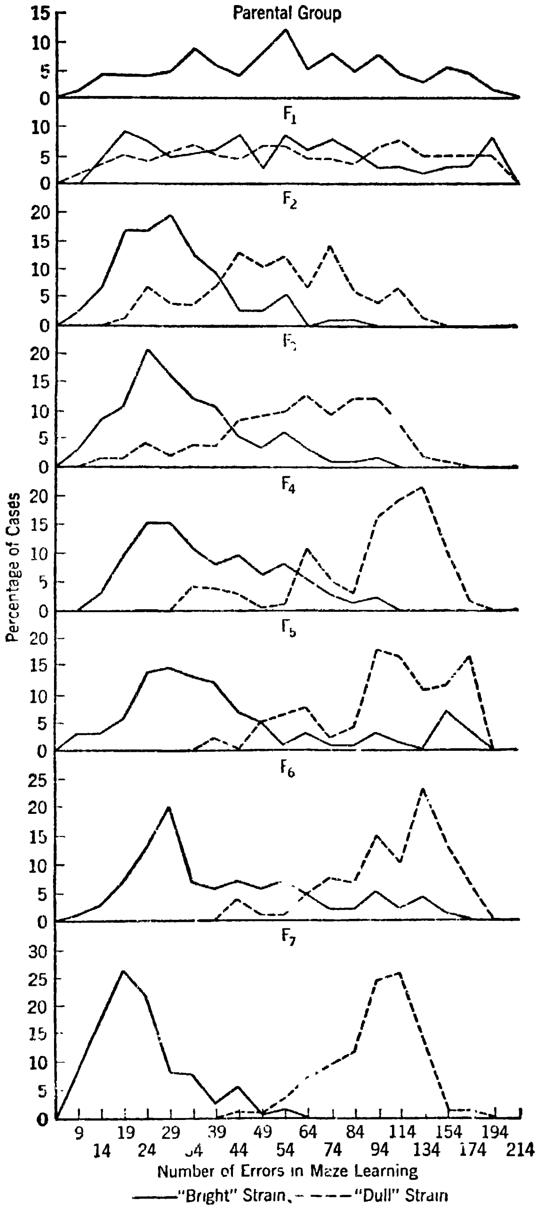


Fig. 37. The Effect of Selective Breeding for Maze Performance. (From Tryon, 53, p. 113.)

One of the most extensive of these investigations employing selective breeding is that conducted by Tryon (51, 52, 53) on maze learning in white rats. An initial group of 142 rats were given 19 trials in running a maze, and the number of "errors," i.e., entrances into blind alleys, was determined for each animal. The group exhibited wide individual differences in maze-learning ability, the total number of blind-alley entrances in 19 trials ranging from 7 to 214. On the basis of these scores, a group of the brightest and a group of the dullest rats were selected for experimental mating. The "bright" rats in this parent generation (P) were mated with each other, and the "dull" were likewise mated together. This procedure was followed through 18 filial generations (F1 to F18). In each successive generation, the "brightest" rats were selected in terms of maze performance and were bred together, the "dullest" being similarly selected and interbred.[2] Environmental conditions, such as food, lighting, temperature, and living quarters, were kept constant for all rats in the different generations.

The effect of such selective breeding upon maze performance is illustrated in Figure 37. The distribution curves indicate the percentage of rats in each group making the number of errors shown on the baseline. It will be noted that the distributions of the bright and dull sub-groups gradually separate until there is virtually no overlapping between them when the F7 generation is reached. Beyond the seventh generation, the additional effects of selective breeding are negligible. In subsequent generations individual differences within the bright and dull groups remain practically unchanged, and the differentiation between the two groups shows no appreciable increase. When rats from the bright and dull groups were interbred, a distribution similar to that of the original parental group resulted, most of the animals now obtaining intermediate scores, with relatively few at the dull and bright extremes. The distributions of the bright and dull parental groups and of two cross-bred filial generations are given in Figure 38. Tryon suggests that the results of this cross-breeding experiment are consistent with the hypothesis of multiple factor inheritance, some of the factors being "dominant for bright performance, some (but fewer) dominant for dull, and some cumulative" in their effect (53, p. 116).

Extensive analyses of the characteristics of the bright and dull

[2] After the F1 generation, a modified "progeny test" was applied in selecting individuals for breeding, the individual being classified not only in terms of his own maze performance, but also on the basis of the performance of his forbears.

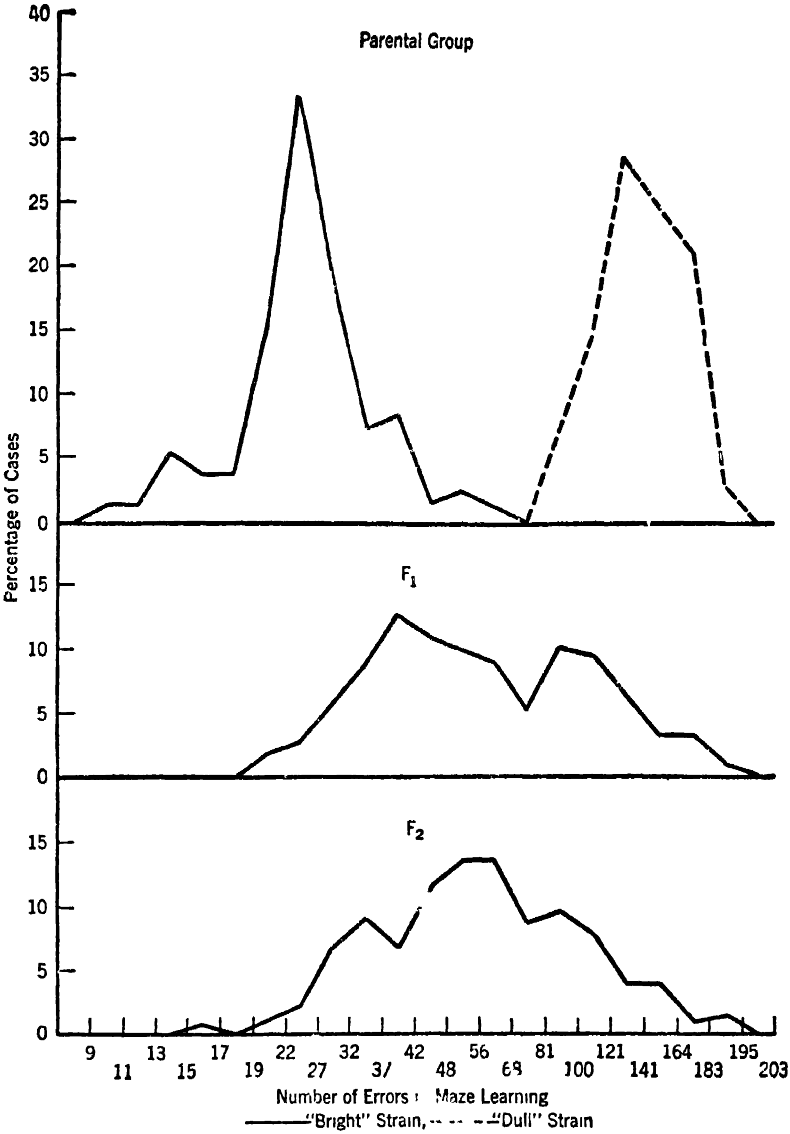


Fig. 38. The Effect of Cross-Breeding Rats from "Bright" and "Dull" Strains. (From Tryon, 53, p. 115.)

groups have demonstrated *consistency* of performance level both when retests were made at later ages and when the error scores, in terms of which the selective breeding was conducted, were compared with other aspects of the animal's maze performance. The evidence does not, however, support the view that the animals differed in *"general learning capacity,"* the difference between the two genetically contrasted groups being specific to the maze situation. Such specificity of learning behavior has been corroborated by other investigators. Experiments designed to disrupt various *sensory cues* available in the maze learning situation showed relatively negligible effects on the performance of the bright rats, many showing no disturbance at all following cue disruption. The performance of the bright rats thus appeared to be largely independent of possible differences in sensory acuity. An analysis of the frequency of errors in each alley showed a consistent difference in the *pattern of error frequencies* between the two groups, a difference which persisted, on the whole, when another maze was employed. Such a finding suggests a possible difference in the animals' approach to the solution of the maze.

Differences in *emotionality* were also found between the contrasted groups, the bright animals showing less emotional response, such as hiding, avoidance, escape, and vocalization reactions, when encountering novel inanimate objects in the maze, but more neurotic behavior in response to handling. The bright group was significantly superior in certain *physical characteristics*, surpassing the dull group in brain size, brain weight, and body weight.[3] The dull group, on the other hand, excelled in fertility. Tryon regards the interpretation of these physical differences equivocal "because of the intensive inbreeding that has occurred during the process of selective breeding" (53, p. 116).

Strains of "maze-bright" and "maze-dull" rats have been bred by Heron (23, 43) through the sixteenth filial generation. In this study, no significant differences in *brain weight* or in ratio of brain weight to body weight were found between the two contrasted groups[4] (43). Differences in *speed of running* (23), however, favored the bright rats. A further analysis (22) indicated that the group which excelled in maze performance manifested a higher *level of general activity* and

[3] Since brain weight is closely correlated with body weight, it is unfortunate that ratios of brain weight to body weight are not given

[4] In fact, most of the comparisons tended to show greater brain size among the dull rats, but the mean differences were all small, the largest being 2.21 times its standard error

stronger *motivation.* Observation of behavior in the maze situation corroborated the hypothesis that such differences operated in the maze learning, the dull rats being described as behaving more like non-hungry rats or hungry rats after the food has been withdrawn from a previously learned maze. These differences were not, however, sufficient to account for the entire difference in maze learning, and other factors very probably contributed to the level of maze performance of the two groups.

Several groups of animals have been directly bred for emotionality and for general activity level. Rundquist (39) selected rats on the basis of *spontaneous activity* as measured by performance on a rotating drum. Selection was carried out through the fifth generation solely on the basis of individual activity; beyond that the extreme individuals within the active and inactive strains, respectively, were bred, with no crossing between the two strains. At the F12 generation, the two strains were well separated in activity level. This investigator reports that activity level showed little, if any, relationship to maze learning. A follow-up study of the same two strains is reported by Brody (3). Selective breeding of these animals until the F29 generation led to a marked decrease in the mean activity, as well as in individual differences in activity level, in the inactive strain, but no change in the active strain. Apparently, active rats were eliminated from the inactive strain, but not inactive rats from the active strain. From the results of strain crosses and back crosses, Brody proposed a specific hypothesis regarding the genetic transmission of the factor determining activity level in rats.[5] The author also points out, however, that environmental conditions seem to obscure the segregation of hereditary factors in some of the cross-matings.

Hall (19) obtained strains of *emotional and unemotional* rats by selective breeding. Subsequent observation of these two strains showed that the emotional rats tend to have a lower activity level in free situations, and to be more variable or less stereotyped in a situation calling for choice. In a later study (20) on the same strains, it was

[5] Specifically, Brody concludes that: ". . . the two strains differ with respect to a single gene rather than with respect to multiple factors. . . . The gene apparently behaves as a dominant in the males and as a recessive in the females. . . . The gene which determines inactivity must act as an inhibitor since none of the matings within the inactive strain produce active offspring, but, on the other hand, active strain matings produce individuals which vary from extreme inactivity to extreme activity" (3, pp. 23 24)

found that the rats from the non-emotional, fearless strain were considerably more aggressive than those from the emotional, timid strain. This result suggests that aggressiveness may be related to genetic factors. An investigation of *fighting behavior of male mice* (40) also showed sharp differences between inbred strains.

In evaluating such studies on the genetic bases of emotional responses, it should be noted that a number of other investigations have suggested the dependence of emotional and "neurotic" behavior in the rat upon certain environmental conditions, such as diet, "taming," and previous relations with cage-mates (cf., e.g., 12, p. 223; 18; 38). To demonstrate that a certain phenomenon depends upon heredity does not, of course, preclude its dependence upon environmental factors, and vice versa.

The researches of Stockard (48) and his collaborators on different breeds of dogs are relevant to the present approach, although in these studies the animals were not bred directly for behavior characteristics. Two groups of dogs were chosen which present a strikingly different picture both structurally and behaviorally. One group, consisting of basset hounds, was characteristically inactive and lethargic; the other, including German shepherd and Saluki dogs, represented the opposite extreme of activity and alertness. Consistent and clear-cut differences in the behavior of these two groups were observed in a series of conditioning experiments. When only members of the same breed were mated, successive generations were found to "breed true" for both morphological and behavior characteristics, i.e., the offspring showed the same characteristic structural and behavioral pattern as the parent generations. Cross-breeding yielded a distribution of behavior types in the F1 and F2 generations consistent with the hypothesis of multiple-factor inheritance.

Differences in endocrine activity between the breeds were suggested as a likely basis for the observed behavior differences. The investigators point out that the basset hound has a relatively inactive thyroid, giving the animal a low metabolism. This condition, together with the correspondingly low activity of the other glands, is probably a factor in the animal's characteristic inactivity and lethargy. The German shepherds and Saluki, on the other hand, have highly active thyroids. Considerable evidence confirming this explanation is furnished by an extensive series of investigations involving the removal of various endocrine glands, as well as the experimental administration of glandular

extracts. The subjects were dogs of known genetic history, mostly from the F2 generation of the above-mentioned crosses. In behavior, they were intermediate between the two extreme types. Through experimental glandular control, their behavior could be made to vary in the direction of either of the two extreme genetic types.

In summary, it may be noted that through selective breeding it is possible to produce strains which are clearly differentiated and sharply contrasted in such behavior characteristics as activity level, emotionality, and maze learning. Some data are also available regarding the structural characteristics which underlie these behavior differences. Glandular conditions, body size, brain size, and factors related to health, vigor, and strength of the hunger drive are suggested as possible bases for the strain differences in learning behavior. Comparisons of the physical characteristics of the contrasted strains sometimes yield inconsistent results, one investigator finding a significant difference in a particular physical characteristic, while another finds no difference in the same characteristic. These inconsistencies are not, however, unexpected if activities such as maze learning are influenced by a *multiplicity of structural conditions*. For example, if we suppose that maze learning can be facilitated by six different structural factors (a, b, c, d, e, and f), an individual experimenter who selects good maze performers in the parental generation may get rats which by chance excel in four of these six relevant structural characteristics (a, b, c, and d). The extensive inbreeding which follows in successive generations will augment *these particular structural differences*, since in effect the experimenter was selecting the animals in terms of these characteristics even though he may have been unaware of it. By the same token, another investigator who singles out his "maze-bright" rats for mating may be selecting them in terms of structural characteristics d, e, and f. In that case, successive generations of selective breeding will produce strains differentiated in d, e, and f, but not in a, b, and c

THE NORMATIVE DEVELOPMENTAL STUDY OF BEHAVIOR

Charting the course of behavior development is of considerable theoretical as well as practical interest in its own right. In the present connection, however, we are concerned only with the use which has been made of such studies in an attack upon the factors which determine behavior differences. In examining the "stream of behavior" as it

appears in the growing organism, investigators have looked for any clues concerning the mechanism of behavior development Thus, the '*sudden emergence*' of functions and the "*sequential patterning*" of development have been commonly regarded as evidence for the importance of maturational factors It has been repeatedly argued that behavior which appears suddenly, in more or less final form, when the organism has reached a certain age, is unlearned The uniformity of developmental stages, or sequential patterning, in any particular function has likewise been cited as a criterion of unlearned behavior, on the grounds that opportunities for learning are likely to vary from one individual to another and could not result in such a consistent succession of like stages We shall consider these criteria in the light of some of the data to be reported below

Observations of the normal course of behavior development have been made extensively on infrahuman as well as human subjects, and during prenatal as well as postnatal development The relative simplicity of the processes in lower forms facilitates the recognition of the essential characteristics of development whose applicability to higher forms can then be more readily studied Similarly some of the most significant observations have been made at the prenatal and neonatal[e] stages partly because of the greater simplicity of behavior at these levels and partly because environmental diversities and opportunities for learning are not so great as in the case of older subjects To be sure investigations on the "growth curve" of psychological functions and on the "growth and decline of intelligence" also logically belong under the heading of normative developmental studies Methodologically, however, such investigations have much more in common with other approaches employing psychological tests and can therefore be more effectively evaluated in connection with the latter (cf Ch 9)

The studies to be discussed in the following section represent only a few outstanding investigations of the behavior development of infrahuman subjects, a field in which a wealth of data is gradually accumulating The available information on human fetal development (to be considered in a later section) is relatively less extensive although it appears to be remarkably full in the light of the methodological difficulties involved in its acquisition In the study of the prenatal behavior of infrahuman organisms, a number of methods are available

[e] The term "neonate" commonly refers to the child between birth and approximately one month of age (cf Pratt 36)

for well-controlled and prolonged observation. In animals such as salamanders and frogs, which pass through a larval stage, the independent-living, immature organism can be directly observed. Bird embryos have been studied through a transparent window made in the shell of the egg. In marsupials, such as the kangaroo, the immature fetus completes its development in the external pouch of the mother, where it is readily visible. Mammals, such as the guinea pig, rat, and cat, have been studied by removing the fetus and keeping it in a physiological saline solution while it is still attached to the mother's body. By this method the fetus may be kept alive long enough to permit relatively extensive observations.

Such experimental procedures are obviously impossible with human subjects. Some scattered information regarding fetal movements can be obtained from introspective reports of the mother, as well as through instrumental observation, as with a stethoscope or recording tambour. The principal source of data on human fetal behavior, however, is furnished by fetuses removed from the mother by Caesarean section, when the health of the mother necessitated such an operation Under these conditions, the fetus has no source of oxygen and can be kept alive for only a relatively short time. It is during this brief period that the behavior observations must therefore be made. Moreover, the behavior of such a fetus will be influenced by the fact that during this time the fetus is gradually dying from lack of oxygen. The probable effect of this condition is initial overactivity followed by underactivity. Finally, the small number of fetuses available for such observations and the unsystematic, uncontrolled variation in their ages further increase the difficulties of obtaining a clear, coordinated picture of behavior development in the human fetus.

The study of postnatal behavior development in *human infants* obviously presents no such methodological difficulties. The obtaining of adequate and representative samplings of subjects for observation at these early ages is, however, considerably more difficult than when children of school age are employed Age records must be precisely reported (preferably in terms of days) owing to the rapid rate of early development. The employment of standardized equipment, including toys, cribs, chairs, stairs, etc., is essential if comparisons among different subjects are to be made. Another important methodological refinement is the use of a one-way-vision screen to eliminate the effects of stimulation by adult observers. Motion pictures have frequently

been employed for detailed and unambiguous recording of behavior. Since so much of the behavior investigated at these age levels is motor. photographic techniques are well adapted for such records. In the development of most of these methodological procedures for the observation of infant behavior, Gesell and his co-workers have made pioneer contributions (cf. 13, 14, 15, 16).

BEHAVIOR DEVELOPMENT IN INFRAHUMAN SUBJECTS

Although a considerable body of data had been accumulated by earlier investigators (cf. 4, 37), the researches conducted by Coghill and his co-workers during the first three decades of the present century represent a major turning point in the study of early behavior development. Coghill's work exerted a profound influence upon both the theory and the experimental study of behavior development. One of the most intensive and best known of Coghill's studies was conducted on embryos of the salamander *Amblystoma* (cf. 7). In his observations of the movements made by this animal in response to stimulation by a fine hair, Coghill noted a uniform succession of stages. At first the animal is non-motile, giving no observable response to the stimulus. The earliest movement is a bending of the head to the right or left. In older embryos, this develops into a bending of the entire trunk, making the animal resemble the letter C. Still later, this C-reaction becomes exaggerated and the animal bends into a tight coil when stimulated. Finally, an S-reaction appears, as a combination of two successive and overlapping C-reactions in reverse directions. Thus, for example, the first C-reaction, toward the left, begins at the head end and travels by progressive muscular contractions toward the tail end; but before this reaction reaches the tail, a second C-reaction, this time toward the right, begins at the head end. When these S-reactions follow each other in rapid succession, their performance exerts pressure upon the water and enables the animal to swim away. Different stages in this behavior sequence are illustrated in Figure 39.

The Coghillian S-reaction, with its characteristic antecedent stages, is held by many investigators to be fundamental in the locomotor development of many animal forms. Not only has it been observed in the swimming movements of numerous aquatic animals, but it also appears in the locomotion of land animals. It can be recognized, for example,

in the motion of the immature opossum or kangaroo as it travels to the mother's pouch immediately after birth (cf. 4).

On the basis of his investigations on *Amblystoma* and other forms, Coghill proposed certain generalizations regarding the sequence of behavior development. Foremost among these is the statement that

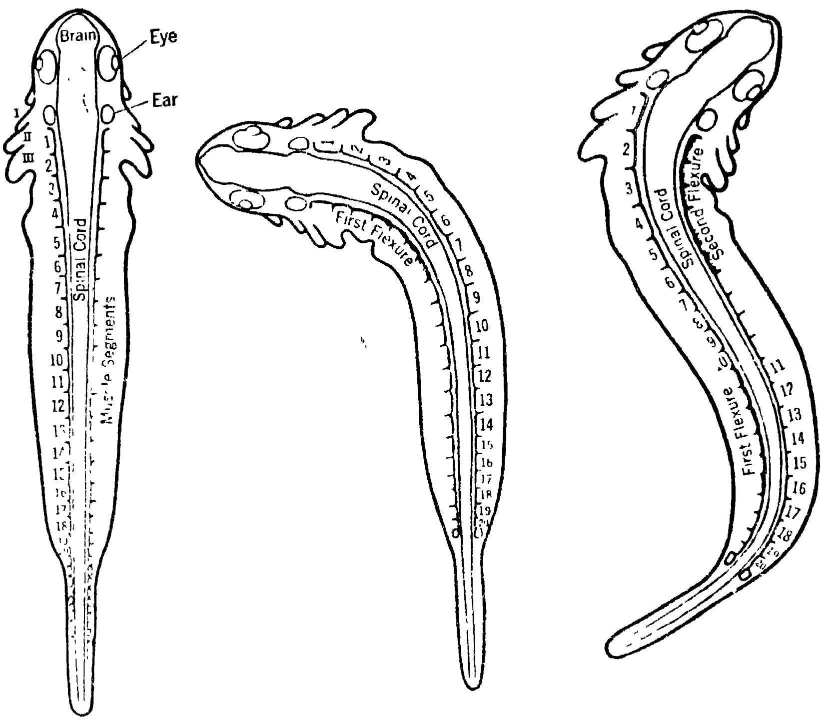


Fig. 39. Successive Stages in the Development of Swimming Movements in the Larva of *Amblystoma*. (From Coghill, 7, pp. 7, 8.)

behavior develops by *individuation* from a total pattern into progressively smaller units. This is virtually the reverse of the view that the earliest acts are simple reflexes through whose combination and integration complex behavior develops. From his observations, Coghill maintained that movements of the whole trunk precede movements of the limbs, the latter being in turn followed by movements of the fingers. Thus, he found "that the first limb movement is an integral part of the total reaction of the animal, and that it is only later that

the limb acquires an individuality of its own in behavior" (7, p. 19). He concludes from such findings that: "Behavior develops from the beginning through the progressive expansion of a perfectly integrated total pattern and the individuation within it of partial patterns which acquire various degrees of discreteness" (7, p. 38). Two additional, related generalizations are that development proceeds along *cephalocaudal* and *proximodistal* axes. The former is based upon the finding that movements of the head region generally appear at an earlier age than movements of the rest of the body, and the progression tends to be from head end to tail end. The latter refers to the succession of development from the trunk outwards; the farther a part is from the trunk, the later, in general, will it exhibit independent movement.

An example of the application of these generalizations to a much higher animal form is furnished by studies of behavioral development in the fetal[7] cat by Coronios (8) and his collaborators. The precise age in days when a large number of different reactions first appeared was noted, together with their subsequent development or disappearance. For example, crawling first occurred in the cat fetus on the 53rd day following fertilization, swallowing on the 51st day; the characteristic "righting reaction" of the cat on the 47th day; and tongue protrusion on the 30th day. Unilateral head bending was first observed on the 23rd day, and the Coghillian C-reaction was found to occur from the 31st to the 45th day of fertilization age. Coronios found evidence of both a cephalocaudal and a proximodistal progression in the behavior development of the fetal cat. His findings also corroborated Coghill's individuation theory, the earlier reactions being relatively diffuse, unorganized movements of the entire organism, and progressing by regular stages to more precise, well-coordinated responses within a narrowly circumscribed area.

Among the other conclusions reached by Coronios on the basis of these studies, special interest attaches to his statements that: "Before birth there is a rapid, progressive, and continuous development of behavior in the fetus of the cat. . . . The 'primitive' reactions of breathing, righting, locomotion, and feeding are the products of a long

[7] In referring to prenatal development, the terms "*germinal*," "*embryonic*," and "*fetal*" are applied to successive stages. In the human, for example, the germinal stage lasts for approximately two weeks after fertilization; from that time until the age of two months the organism is known as an embryo, and from two months until birth, as a fetus. The duration of these stages, of course, varies with different species.

and continuously progressive course of prenatal development" (8, pp. 377–378). A consideration of the extent to which behavior development occurs before birth, as illustrated by these observations of Coronios, is important for the proper evaluation of some of the observations made on postnatal behavior development. For example, in earlier studies by Tilney and his associates (50) on the behavior development of the cat after birth, emphasis was placed upon the "sudden emergence" of a number of reactions at specific postnatal ages. Several writers have cited these results as strong support for the role of maturation in behavior development. All that such observations may show, however, is the sudden reappearance, in the presence of the suitable environmental stimulus, of behavior which has undergone a more gradual development during the long prenatal period.

The role of prenatal environment in the development of behavior is stressed by Kuo (28, 29, 30, 31) on the basis of his extensive and carefully controlled studies of the chick embryo. The observation of the first appearance of different responses tended again to corroborate Coghill's hypothesis that development of behavior follows cephalocaudal and proximodistal progression. Kuo reports that "head movements appear first, trunk movements next, and those of the extremities and tail last" (28, p. 406).

Besides charting the order of appearence of different reactions, Kuo investigated the possible contributions which mechanical and other environmental factors operating during prenatal life might make to behavior development. He noted, for example, that the beating of the heart early in prenatal development produces a general rhythmic vibration of the inert fetal body, which starts the head on passive mechanical movement. Similarly, the mechanical movement of the fetus by the contractions of the amnion [8] stimulates the fetus to make active movements. A marked increase in active fetal movements is found at the period of greatest amnion activity. Eye reflexes were likewise observed before birth, despite the absence of visual stimuli; such reflexes occurred in conjunction with movements of the body. From observations such as these, Kuo concludes that every part of the muscular response system of the chick has been exercised before birth, and that many organs begin to function when still in a rudimentary form. He lays considerable emphasis upon the fact that the development of behavior

[8] The sac in which the fetus is enclosed.

is gradual and continuous, pointing out that so-called instinctive patterns of behavior do not appear suddenly, but have a long developmental history.

Our knowledge of prenatal behavior has been considerably advanced by the thorough and detailed studies of the fetal guinea pig conducted by Carmichael and his co-workers (4, 5). The time of first occurrence as well as the nature of the many responses in this animal's prenatal behavior repertory was determined through carefully controlled studies. In addition, a number of more specific investigations dealt with such problems as the response to temperature and to pressure stimulation of varying intensity. The latter studies indicated the influence of stimulus intensity upon the nature of the response, a light pressure on a particular point of the skin, for example, eliciting only an eye wink, while a more intense pressure on the same point led to

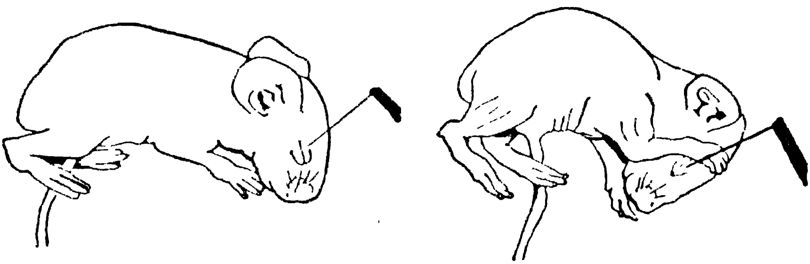


Fig. 40. Contrasting Reactions of the Fetal Guinea Pig to Light and Heavy Pressure Stimulation. (From Carmichael and Smith, 6, p. 432.)

movements involving head, entire trunk, and all four limbs. These two responses are shown in Figure 40. It should be noted in connection with Coghill's generalizations that the more intense stimulation tended to evoke a more general response, while weaker stimuli called forth more specifically localized reactions, *regardless of the age of the fetus.*

Carmichael's results on the whole suggest the need for qualifying any generalizations that had previously been proposed regarding the course of behavior development. Thus he found that the earliest behavior of the guinea pig includes some responses which are "generalized" in Coghill's sense, but also some which are highly specific and narrowly localized. Carmichael suggests that many of the early responses which appear to be general may actually be series of responses

to successive proprioceptor [9] stimulation. Such stimulation, spreading through the organism, may lead to responses which may easily be mistaken for diffuse and undifferentiated behavior. Carmichael's studies also revealed a number of exceptions to the cephalocaudal and proximodistal progression of behavior development, indicating that these generalizations, too, need to be qualified.

PRENATAL BEHAVIOR DEVELOPMENT IN HUMAN SUBJECTS

The major studies of human fetal behavior reported to date are those by Minkowski (34), Bolaffio and Artom (2), Hooker (25, 26), and Sontag (45, 46). In Hooker's investigations, moving-picture records supplemented the stenographic notes of the earlier workers, thus making possible later, detailed analyses of responses. Hooker's *Preliminary Atlas of Early Human Fetal Activity* (25), based upon these photographic records, is a valuable source of data on human fetal behavior. The research on fetal behavior being currently conducted by Sontag and his associates at the Fels Research Institute, as a part of the developmental studies of the Institute, has already furnished promising results on a number of questions.

The available evidence indicates that all receptors are probably capable of functioning in the fetus, although the conditions of prenatal life are such that vision, taste, smell, and temperature are not likely to be stimulated. Since change is a primary requisite for stimulation, the uniformity of the prenatal environment in such characteristics as temperature and chemical composition makes the activation of certain receptors unlikely. Responses to auditory stimulation *in utero* have, however, been noted. Touch stimulation has been extensively studied and found to occur shortly after the eighth week. Although a general cephalocaudal and proximodistal succession may be observed in the development of such sensitivity, exceptions have again been noted. Proprioceptors probably begin to function at the time when the first active movements are made, when the fetus is about two months of age. The importance of such stimulation in producing what appear to be generalized movements of the entire organism is again emphasized by Carmichael, in his survey of the available data on fetal behavior.

[9] The proprioceptors are located in the muscles, tendons, and joints, and furnish stimulation to the organism from its own movements.

In reference to organic senses in the fetus, Carmichael writes: ". . . it may be said that there are possibly certain organic changes in the stomach, intestines, heart, and vascular and respiratory systems which occur before birth and which may be important in receptor stimulation" (4, p. 131).

Thus it is apparent that the fetus has various sources of stimulation in its normal prenatal environment. Stimulation may arise from internal changes, as well as from tactual and proprioceptive impulses resulting from movements of the fetus itself. The "spontaneous" movements noted in the fetus probably represent responses to such stimuli, unrecognized by the observer. Mention may also be made of the possibility that the passive movement of the fetus by the contractions of the amnion, occurring after the third month, may in turn stimulate active motor responses, as was suggested by Kuo for the chick embryo (cf. 4, p. 104).

Following the first two months of prenatal life, a variety of motor responses have been observed. While many of these responses tend to involve the entire organism, a number of local reflexes also appear early in the course of development. By the fourth month, nearly all the reflexes of the newborn can be elicited. It is interesting to note that a number of reactions whose proper stimuli are not present until after birth are nevertheless found during prenatal life. Thus crying, sucking, and eye reflexes, for example, have been observed in the fetus. Among the early movements of the fetus are rhythmic contractions of the chest and thorax, similar to breathing movements, which can be recorded directly through the mother's body. Since the fetus is suspended in a liquid, true lung breathing is, of course, impossible; but the neuromuscular mechanism of breathing may be exercised and strengthened by these movements (cf. 4, p. 104).

A number of fetal responses appear also to be the precursors of later postural and locomotor behavior of the infant. Reflex "balancing" responses, for example, have been observed in the fifth month of prenatal life. At this time, movements of the head in space lead to equilibratory movements of the limbs. Passive movement of the fetus produces active reactions which return the fetus to its original position. Stimulation of one foot in a five-month fetus may lead to the bending of the corresponding leg and the extension of the opposite one; a diagonal response of the opposite hand may also result. This "trot" reflex may underlie such postnatal activities as crawling and walking.

An increase in the amount and rate of fetal movement has been observed during emotional excitement of the mother (45). Apparently the physiological changes occurring in the mother's body during emotional stress excite the fetus to greater activity. There is also some evidence that hyperactivity, excitability, and feeding difficulties may occur for several months in infants whose mothers have undergone severe emotional strain during pregnancy (45). The physiological conditions of prenatal life associated with the mother's emotional state may thus have effects which continue in postnatal life.

Can *learning* occur during prenatal life? The question is a controversial one, but there is evidence that at least simple modifications of behavior do occur following stimulation. Adaptation to a vibratory stimulus applied to the mother's abdomen has, for example, been observed in the fetus. Ordinarily such a stimulus will evoke a typical "startle reflex" involving sharp, convulsive movements and a sudden rise in heartbeat (45, 46). A definite decline in this startle response was noted in a fetus repeatedly stimulated in this manner over several weeks (45). Rudimentary conditioning has also been obtained in human fetuses *in utero* during the last two gestation months (47).[10] In these experiments, a loud noise served as the unconditioned stimulus, and a vibratory stimulus applied to the mother's abdomen as the conditioned stimulus. Approximately 15 to 20 paired stimulations were required to establish the conditioned reaction to the point at which 3 or 4 successive responses to the vibratory stimulus alone were obtained, while additional reinforcements led to as many as 11 successive conditioned reactions. Experimental extinction, spontaneous recovery, and retention of the response over a three-week interval were also demonstrated. Several investigators (cf. 35, p. 209) have succeeded in establishing conditioned reactions in the *neonate* to a variety of stimuli. The suggestion that some of the responses of the fetus and the neonate may be conditioned reactions has been made by Holt (24) on theoretical grounds.

BEHAVIOR DEVELOPMENT IN HUMAN INFANTS

Mention has already been made (cf. Ch. 2) of the detailed *normative scales* which chart the normal course of postnatal behavior develop-

[10] It may also be noted that Gos (17) reports conditioning of the chick before it has emerged from the egg.

ment in the human infant and young child. The investigations of Gesell (13, 14, 15, 16) and his co-workers at the Yale University Clinic of Child Development are probably the most extensive and well controlled of such normative developmental studies. Since the behavior repertory of the child during the first few years consists so largely of simple sensory and motor functions, most of the data concern this type of activity.

Gesell concludes from various findings that the behavior development of the infant and child depends primarily upon "growth" or "maturational" factors, rather than upon learning. As one source of evidence, he cites observations on *pre-term* and *post-term babies*. Infants born one month before the normal nine-month gestation period do not reach the developmental level of the normal newborn child until the age of one month. Similarly, a baby born after a ten-month gestation period will be as far advanced in its behavior at birth as a normal one-month-old child. Yet, Gesell points out, there is a vast difference between the prenatal and postnatal environments in the opportunity for learning and for the specific exercise of behavior functions.

Gesell likewise points to the *consistency of developmental sequences* as evidence for maturation. In the development of prehension behavior, for example, the successive stages follow in the same order and at approximately the same ages in different children. Thus the child's reactions toward a small sugar pellet placed in front of him show a characteristic chronological sequence in visual fixation and in hand and finger movements. Use of the entire hand in crude attempts at "palmar prehension," for example, occurs at an earlier age than the use of the thumb in opposition to the palm; this is in turn followed by the use of the thumb and index finger in a more efficient "pincer-like" grasp of the pellet. Such sequential patterning is likewise reported for walking, stair climbing, and most of the sensori-motor development of the first few years.

A similar emphasis upon maturation is to be found in other studies of behavior development in infancy (cf. 1, 9, 11, 21, 33). Shirley (41, 42), in a study based upon weekly and biweekly examinations of 25 infants from birth to two years, concludes that "motor control in infancy begins headward and travels footward," thus supporting the principle of cephalocaudal sequence. In a summary of her observations, she writes:

The first stage of the motor sequence was eye-coordination, which had as sub-stages furtive pursuit movements, fixation of persons and objects, and following moving objects. During the same interval postural control moved down the body from head-turning and head-lifting to chest-lifting in the prone position. Likewise there was gradually less need for support of head, neck, and back when the baby was held on the shoulder. The second period saw advancement in postural control and development of reaching. When the baby was seated on the examiner's lap her hands gave him support first at arm-pits, then at mid-ribs, abdomen, and finally none save the cupping of her lap. Reaching progressed from random waving toward the object through touching, grasping, retaining with thumb opposition, to the goal of carrying object to mouth. With the advent of sitting alone motor control had migrated down to the sacral region, and with the use of the index finger for pointing, which occurred about the same age, it had crept out to the finger tip (41, p. 204).

Shirley's strong leanings toward a hereditary interpretation are exemplified in the following quotation:

Can an order which holds so universally be attributed to training or even to spontaneous practice? Does not its conformity to the anterior-posterior growth law make the motor sequence a normal unfolding [sic!] of developmental processes—in other words, a function of maturation? (41, p. 205).

In evaluating the conclusions reached by such investigators, the following points ought to be borne in mind. First, the possibilities for *prenatal exercise* of various functions as well as for *prenatal learning*, suggested in the preceding discussion, should be noted. Dennis (10) has argued against such a view. Pointing out that neither operant conditioning[11] nor trial-and-error learning has been demonstrated in the human fetus or neonate, he concludes that most of the behavior of the infant at birth is unlearned and that fetal development is almost entirely a matter of maturation. The question is certainly not yet settled, however, and requires more than the relatively meager information now available on human prenatal development. It should be noted, furthermore, that the simple exercise of various functions, initiated and determined by the physical conditions of the prenatal environment, might still influence subsequent behavior development,

[11] Operant conditioning involves the conditioning of random, spontaneous, or self-initiated acts, and is thus the counterpart in conditioned-response terminology, of trial-and-error learning. For a more technical discussion of the concepts of operant and respondent behavior, cf. Skinner (44).

even if operant conditioning and trial-and-error activity are shown to be impossible in the fetus.

In the second place, the importance of guarding against easy generalizations must again be emphasized. Superficially, infant behavior may appear to fit the cephalocaudal and proximodistal sequences, but, as in the case of infrahuman and prenatal development, *exceptions can be found.*

Thirdly, as applied to certain forms of behavior, these generalizations seem to be the result of the *physical exigencies of the situation* rather than the result of a developmental progression in the locus of motor control. Let us consider, for example, the observation that the baby, in order to sit up, requires support first at the armpits, then at mid-ribs, then at the abdomen, and finally none besides the cupping of the holder's lap. Now, if the child is supported at the abdomen, his own muscles have to support the upper half of the body. If, however, he is supported at the armpits, the intervening parts of the body are thereby supported as well. Thus the "armpit support" is physically more complete than the "abdominal support." The former would naturally be required with younger children whose motor control is still inadequate. The same may be said of the fact that sitting up precedes standing. The child has less to support through his own muscular efforts when seated than when standing. Similarly, it could be argued that finger movements, in order to be used effectively, require more delicate and finer adjustments than do the grosser movements of hands and arms, and may appear later simply because they require more highly developed muscular control. Certainly no special law of "inner growth" is needed to account for the fact that movements requiring the coordination of more muscle groups or those requiring finer coordination appear later than movements requiring fewer muscles, less strength, or less delicate coordination. Learning ordinarily progresses from the simpler, easier aspects of a task to the more complex, and would thus be completely consistent with the type of progression observed in these developmental studies.

A fourth point to be noted is that uniformities of developmental sequences may result in part from certain basic *environmental similarities* in the average American homes in which the children have been developing. Not only deliberate instruction on the part of adults, but also other, unplanned uniformities in the child's physical and psychological milieu need to be considered. Under these conditions, a regu-

larity of sequence need not necessarily imply maturation. Finally, it should be remembered that the behavior under investigation, being largely of a *simple sensori-motor nature,* would naturally depend upon the level of structural maturity attained at the time. It would be unwarranted to generalize from such observations to the development of more complex, symbolical activities in the older individual.

STRUCTURAL CORRELATES OF BEHAVIOR DEVELOPMENT

Any contribution which hereditary factors may make to behavior development must obviously operate through structural characteristics, which would in turn set certain limiting conditions to the acquisition of behavior. As was pointed out in Chapter 4, this is the only way in which such hereditary influences may be manifested. To study what structural changes occur concomitantly with the observed developmental changes in behavior thus constitutes another approach to the heredity-environment problem.

Coghill (7) was the first to trace systematically the role of specific *anatomical relationships within the nervous system* in the development of a particular behavior function. In his studies of the salamander larva, Coghill noted that while the animal is still in a non-motile stage, there are both sensory and motor nerve fibers in contact with receptors and muscles, respectively, but there is no central connection between the two. At this stage, the animal does not respond to tactual or chemical stimulation of the skin, however intense. Regarding subsequent development, Coghill writes:

> With the ability to respond to tactile or chemical stimulation of the skin there appears a third series of cells. They bridge the gap between the sensory system of one side and the motor system of the other. Their bodies lie in the floor plate of the medulla oblongata and upper part of the spinal cord. In the non-motile stage these cells are unipolar. The one pole of the cells extends either to the right or to the left into close relation with the motor tract on one side only. When they become bipolar, they complete the path from the sensory field to the muscle . . . (7, pp. 12–13).

The presence of these bipolar cells is sufficient for the development of muscular responses through the "coil" stage described in the preceding section. Additional neural structures must appear, however, before swimming movements can occur. In summarizing this stage, Coghill writes:

> At the time swimming begins there is a growth of collaterals from fibers of the anterior part of the motor tract into relation with the dendrites of floor plate cells. These collaterals cause an excitation that is on its way to the muscles of one side to be carried through the commissural cells of the floor plate to the motor system of the other side. But in this passage to the muscles of the opposite side more synapses are involved than there are in the path to the muscles of the same side; so that the second flexure follows the first by a very brief interval. In the same manner as the impulse for the first flexure excited the second flexure so the impulse for the second excites the third, and so on (7, p. 14).

Thus we see how the delicately timed succession of excitations from side to side could serve as a basis for the overlapping right and left C-turns of the body which constitute the swimming movement of this organism. Coghill charted similar relationships between the anatomical growth of the nervous system and the appearance of other behavior functions, such as feeding behavior.

A similar, but usually much less complete, neural basis has been worked out for the earliest behavior development in a number of other animal forms, such as the rat. Through anatomical and histological studies, the development of sensory structures has likewise been related to the appearance of corresponding behavior functions. Some information regarding the neural structures underlying early human behavior development has been obtained from histological studies of human fetuses after death. Observations of non-motile fetuses, for example, have shown that in such cases sensori-motor connections had not yet been established in the nervous system.

All the investigators whose observations of human fetal behavior were cited in the preceding section have also conducted experiments designed to determine the functioning of different levels of the nervous system at various prenatal stages. *Extirpation of the cerebral hemispheres* has been found to have no effect on the motor responses of the early fetus. Until about the end of the third month, the cerebral cortex appears not to function in behavior. Even after three months of fetal age, the effect of cortical removal is slight and not observable in all cases. The only observed effect of decortication at any time during prenatal life is an increase in the intensity of local reflexes following cortical removal, the cortex having presumably exerted an inhibitory influence. Sectioning of the spinal cord in a motile fetus, on the other hand, abolishes certain reflexes in the corresponding body areas. Re-

moval of the cord leads to complete cessation of sensori-motor responses.

Electrical stimulation of the cortex generally produces no response in the human fetus, although such stimulation applied to lower brain centers or to the spinal cord produces the appropriate muscular contractions. In one investigation (2), some evidence was presented that stimulation of the cortex during the last prenatal month arouses activity. Studies of the *electrical brain potentials* of the human fetus have recently been made by means of electroencephalograms obtained *in utero* (32). These also show little or no evidence of brain activity before birth. It is generally believed, from a variety of evidence, that cortical control of behavior in the human does not begin until some time after birth.

A study of the brain potentials of the fetal guinea pig (27), however, gave evidence of electrical activity beginning at about two weeks before normal birth time. In slightly older fetuses, the effects of stimulation could be recognized by definite, though not invariable, changes in the electroencephalogram. The authors concluded that: "The guinea pig brain first exhibits electrical activity at a time when behavioral indications also point to maturation of higher nervous centers" (27, p. 71).

Another attempted approach to the identification of the structural correlates of behavior was based upon the *myelination* of the nerve fibers. Before impulses can be conducted along discrete paths from receptor to nervous system and from there to the muscles, the nerve fibers must be insulated. It was suggested that the myelin sheath which surrounds the nerve fiber provides such insulation and that the appearance of the myelin covering would thus serve as an indicator of the time when different nerve paths begin to function. Tilney and Casamajor (49) applied this method to their studies of behavior development in the newborn cat and guinea pig. Although these investigators reported that the specific reactions which they observed did not appear until the corresponding nerve fibers had become myelinated, these findings were not corroborated by other workers with the same or other animal forms. It is now generally recognized that myelination is not a prerequisite for the functioning of nerve fibers.

Several other factors have been proposed as essential antecedents for sensori-motor functioning. It has been suggested, for example, that the real carriers of the nerve impulse are the *neurofibrils*, tiny

threadlike processes which run through nerve tissue. If this is the case, then the onset of functioning of particular nerve paths could be determined by observing the first appearance of the neurofibrils. Some investigators have noted relationships between the presence of the neurofibrils and behavior development, but the role of the neurofibrils is still insufficiently established for the proper interpretation of these findings. Since chemical changes occur during neural functioning, the *presence of certain chemical substances* has been regarded by some as a possible antecedent of such functioning. In his studies on the chick embryo, Kuo (31), for example, found that the first true neurally mediated responses do not occur until after the presence of acetylcholine can be detected.

This brief examination of typical procedures and findings should serve to characterize the present state of knowledge regarding the structural correlates of behavior development. The clear-cut identification of structural changes underlying the appearance and development of particular behavior functions is rare. Conclusive results have been largely restricted to simple behavior in relatively low animal forms. Much of the available information, especially as it pertains to the human, is highly tentative, exploratory, and sketchy. Certainly the available data furnish no justification for many of the statements frequently made regarding the neural basis of complex behavior characteristics in man. Such speculations may possibly be of some value in suggesting problems for research. But it is important that their speculative nature be clearly recognized.

CHAPTER

6

Psychological Factors in Simple Behavior Development

THE PRESENT CHAPTER is concerned with studies specifically designed to determine the variations in behavior development resulting from changes in environmental conditions. The investigations to be considered are similar to those treated in the preceding chapter in that they deal with relatively simple behavior. Moreover, the organisms studied are themselves either at a relatively low level in the phyletic scale or are studied at an early stage in their development. In the majority of these investigations, environmental conditions were systematically varied by the experimenter and the corresponding changes in behavior were observed. This is true of all the studies on *infrahuman subjects*, to be discussed in the first section. The second section deals with investigations of human infants by the method of *co-twin control*. In such a method, environmental conditions are altered by experimentally providing additional training for one member of each pair of twins. In the third section of the present chapter we shall consider the method of *experimentally restricting the training* of human infants, a method of which relatively little use has been made.[1]

Although not, strictly speaking, experimental, the two remaining groups of studies covered in the present chapter are closely allied to the methods described above. The observation of the behavioral effects of *infant-rearing practices in different cultures*, as well as the case reports of *"feral man"* which have appeared from time to time, may be regarded as relatively crude, unplanned "experiments" on the effects of environmental variation.

[1] The three methods discussed above logically fall under category 4 in our listing of the approaches to the problem of heredity and environment, given in Chapter 4. For surveys of such studies, cf. 7, 32. The other two types of study to be considered in the present chapter fall under categories 5 and 6, respectively.

EXPERIMENTALLY PRODUCED CHANGES IN ANIMAL BEHAVIOR

A number of behavior functions commonly regarded as unlearned or "instinctive" have been subjected to experimental control, whereby the animal was prevented from exercising the function until well past the age when such a function normally appears in the species. By such isolation of factors, the attempt is made to determine the extent to which the physical maturation of the necessary structures will in itself lead to the performance of the given function. Other experimenters have followed the opposite procedure of providing additional intensive training in order to determine how far normal behavior development can be accelerated.

One of the pioneer experiments on maturation *versus* learning was that of Shepard and Breed (34) on the pecking behavior of chicks. One of these authors (Breed, 3) had previously observed that the pecking response involved three separate reactions—striking, seizing, and swallowing—in each of which an error could be made. Defining a successful response as one in which all three processes were completed, he found that, during the first month after hatching, the accuracy of chicks in pecking at grains normally rises from about 15% successful trials to about 84%. Is such improvement the result of the intervening practice or does it follow from the structural development of the sensory, motor, and neural mechanism? To answer this question, Shepard and Breed (34) kept three groups of newly hatched chicks in darkness for periods of three, four, and five days, respectively. During this time the chicks were fed by hand, the experimenter conveying food and water directly into their mouths. Because of the darkness, the chicks had no opportunity to peck at any objects. In terms of the above-mentioned criterion of successful responses, the older chicks—whose pecking had been delayed longer—did no better *initially* than those which had started to peck at a younger age. The older chicks, however, progressed more rapidly. In other words, the groups that had been delayed for several days required *fewer days of practice* to reach the maximum level of accuracy than did the chicks which had had unrestricted opportunity to peck from birth. The chicks which had reached a more advanced stage of physical development thus profited more from practice than did the younger chicks. These results seem

to suggest that, although some practice is necessary, the pecking response also depends upon maturation.

Although widely quoted, the results of this early experiment were inconclusive for several reasons, not the least of which were the small number of cases and the wide variability of performance among individual chicks. More recently, the experiment has been repeated by several investigators, with certain methodological modifications. The most carefully controlled and thorough analysis is to be found in the investigation by Cruze (10). A total of 202 chicks from comparable stock were separated into eight experimental groups, each containing 25 or 26 chicks. The first five groups, A to E, were kept in darkness for 1, 2, 3, 4, and 5 days, respectively. This part of the experiment thus represents a repetition of Shepard and Breed's investigation, under more carefully controlled conditions.

The results were also similar to those of the earlier investigators, as shown in Figure 41. It will be noted that, although each group is about equally poor in initial performance and all show marked improvement with practice, *this improvement is more rapid for the older chicks*. Thus the final level of 20 out of 25 successful trials was reached after 15 days of practice by Group A, which began to peck after only one day of confinement in the dark, i.e., when the chicks were only two days old. At the other extreme, Group E, confined for 5 days, reached the same level of accuracy after only 7 days of practice.

Cruze did not, however, stop at this point, but set out to control one more factor, viz., the *amount of daily practice* subsequent to removal from the dark room. Groups A to E had been free to peck, and did in fact peck extensively, outside of the daily test periods. In other words, although practice had been prevented during the period of dark-confinement, no control was exerted over the amount of subsequent practice. Three additional groups, F, G, and H, were kept in darkness at all times except during the daily tests. All three groups started the pecking tests after one initial day in darkness and are therefore comparable in this respect to Group A. The results obtained with each of these four groups are given in Figure 42. Group A, with 25 trials a day plus unlimited outside practice, quickly outstripped the other three groups in performance. The practice undergone by Group F was restricted to the 25 daily test trials, and its progress is correspondingly slower. Group G, allowed only 12 daily test trials, is

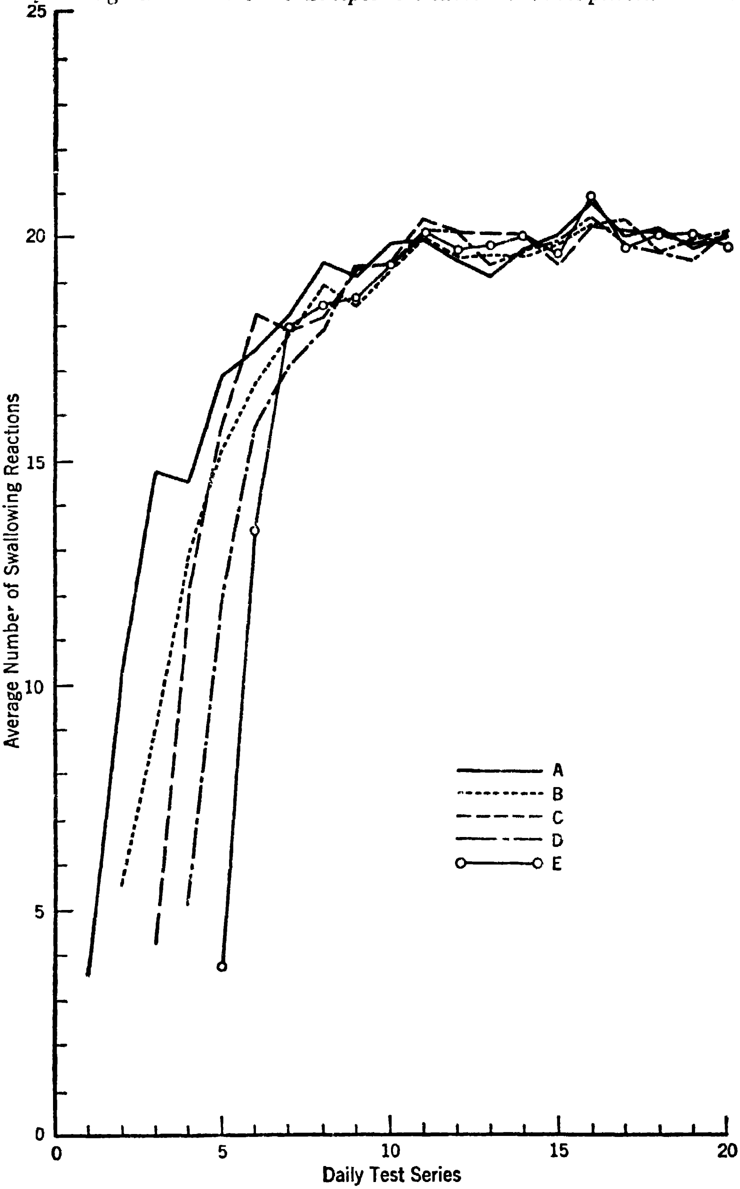


Fig. 41. Pecking Performance of Five Groups of Chicks Prevented from Pecking for 1, 2, 3, 4, and 5 Days, Respectively. (From Cruze, 10, p. 386.)

clearly at the bottom of the groups. In fact, this group shows an almost horizontal curve, with only slight improvement during the first few days. An interesting variation was introduced in Group H, which had 12 daily test trials for 10 days, followed by 25 daily test trials during the next 10 days. The sharp rise in the curve of this group after the tenth day clearly reflects this change in amount of practice. During the first 10 days, progress in this group had been very slow, the curve coinciding closely with that of Group G.

It is apparent from this part of the experiment that extensive practice is essential for reaching a high level of attainment in the pecking response. Only a slight initial improvement could be directly attributed to maturation. Further analysis of the nature of the errors, moreover, demonstrated that the role of maturation is restricted primarily to the development of the striking response. The swallowing response is a reflex which is present before hatching and shows little change in any of the groups. The accuracy of the seizing response was found to depend closely upon practice; this aspect of the pecking function evidently accounts in large part for the effects of practice noted in the case of the total function.

In a comparison of the present investigation with the earlier work of Shepard and Breed, two points are noteworthy. First, with more refined methodology, the important role of learning becomes apparent in the performance of a function which superficially appeared to depend more largely upon maturation. Secondly, the analysis of a complex function may show some of its component activities to be largely the result of learning, others largely the result of maturation. Any attempt to characterize the function as a whole in terms of learning or maturation would thus be misleading.

In a series of investigations by Carmichael (4, 5, 6), the role of maturation and learning in the swimming of tadpoles was investigated by a similar method. While the newly hatched animals were still in a non-motile stage, they were separated into two groups: one of these, the control group, was allowed to develop in ordinary tap water; the other group was kept in a weak solution of chloretone. Although not interfering with neuromuscular growth, the chloretone produced complete immobility, thus making practice impossible. When the control group was swimming normally, the drugged tadpoles (who had reached the same stage of structural development) were removed to fresh water. Within 30 minutes, these tadpoles are reported to have

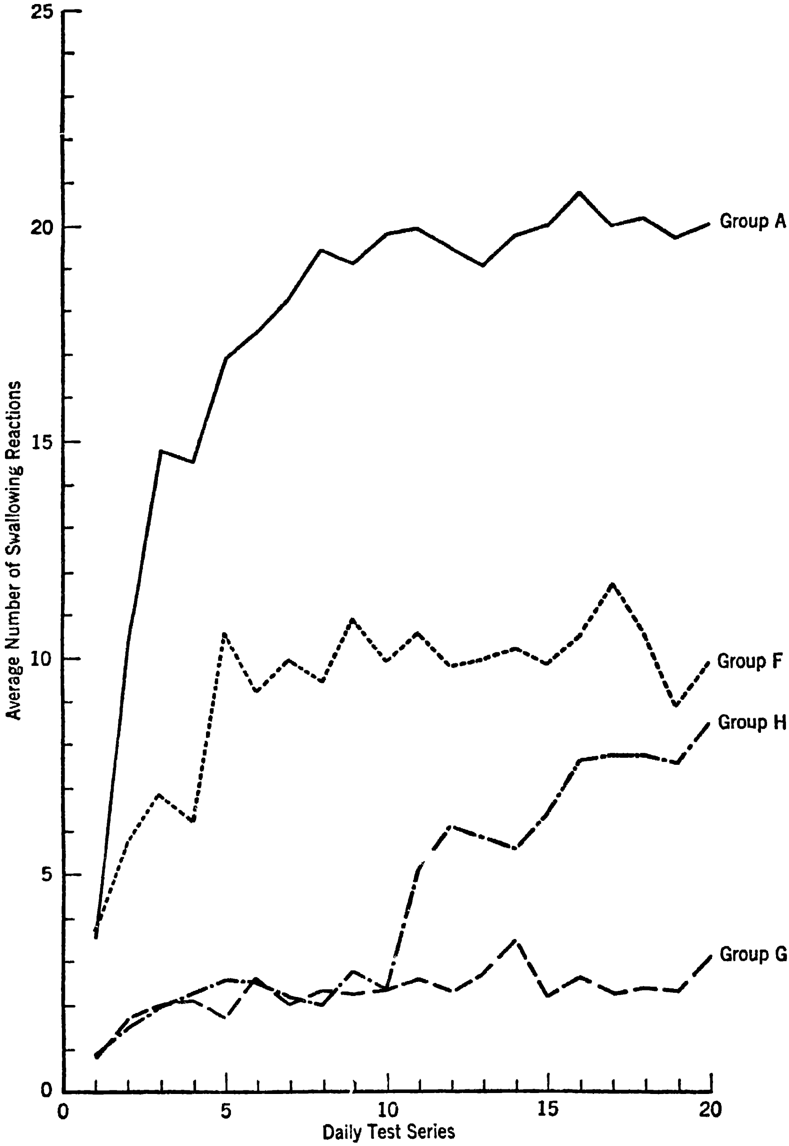


Fig. 42. Pecking Performance of Four Groups of Chicks Allowed Varying Amounts of Daily Practice. (From Cruze, 10, p. 388.)

been swimming as well as the control group, which had had five days of swimming practice.

In a further experiment, Carmichael (5) demonstrated that the 30-minute delay represented simply the time required for the effects of the drug to wear off, and could not be regarded as a period during which rapid learning might have occurred. In still another experiment (6), in which isolation from external stimulation was employed instead of anaesthetization, essentially the same results were obtained. From these investigations, Carmichael concluded that the swimming of tadpoles can be classified as unlearned behavior, depending solely upon structural development.

More recently, Fromme (17) repeated this investigation with a number of modifications. Chloretone was again used to induce immobility, the drugged and control tadpoles being carefully equated in terms of their precise stage of morphological development. A quantitative measure of the velocity of swimming was obtained by testing the animals in a narrow trough which permitted only straight-line swimming. With this measure, the investigator was able to show that, although all the animals were able to swim shortly after removal from the drugged water, significant differences in speed and distance of swimming were present between the drugged and control groups.

Moreover, the developmental stage at which the animals were released from the drug influenced the results. Thus one group, released at the stage when the first movement normally occurs, showed no difference in swimming ability from the control group. This is to be expected, of course, if the only effect of the drug is to produce immobility. A second group was allowed to develop in ordinary water until the appearance of the elementary movements which normally precede swimming. Then it was placed in the drugged water, where it remained until it reached the stage of structural development characteristic of the free-swimming tadpoles. When released and tested at this time, the group swam more poorly than the control group.[2] Finally, a group anaesthetized throughout its development and released and tested at the stage when swimming normally occurs showed an even greater inferiority, when compared with the control group.[3]

[2] The difference between experimental and control groups was statistically significant, the critical ratios being 5.2 for speed of swimming and 5.3 for distance. (For an explanation of statistical tests of significance, cf. Ch. 18).

[3] The critical ratios of these differences were 9.5 and 7.1 for speed and distance measures, respectively.

From all these results, Fromme concludes that, "although the development of structure may be explained completely in terms of the growth process, the development of behavior is determined only in part by the structures produced by growth and is affected by its own behavioral antecedents as well" (17, p. 235).[4]

Similar experiments have been conducted on such activities as the flying and singing of various species of birds, and the behavior of cats in catching and killing mice. The role of maturation and learning in the development of sexual behavior has also been investigated in a number of animal forms. In general, such studies have indicated that some sexually determined activity occurs when certain developmental stages are reached, as a result of endocrine secretions and other physiological factors. The specific way in which such activity is expressed and the object toward which it is directed, however, vary according to environmental circumstances.

In an experiment on male doves reared in isolation from other members of the species, a number of sexual "abnormalities" were observed (9). The birds would bow and coo to the experimenter as normal birds do to members of their own species. They seemed to pay especial attention to the experimenter's hand, with which they came into contact when fed; one bird actually went through the act of copulation while on the hand taking food. Female doves reared in isolation developed similar anomalies of behavior (8). If the experimenter stroked them and preened the feathers of their head and neck, they exhibited characteristic courting behavior. Egg-laying was even induced in many instances by this method. Experimental "homosexuality" was produced in a large number of cases when two female pigeons were reared together. In such cases, the animals displayed the usual courting performance toward each other, followed by egg-laying on the part of both animals.

Equally pronounced variations of behavior were noted in a young monkey separated from its mother at the age of three days and brought up in isolation from all members of the species during the first 18 months of life (15, 16). The development of sexual behavior in gen-

[4] Further experiments on the effects of *additional stimulation*, in which the animals were kept in constantly agitated water, yielded inconclusive results. Several possible reasons may account for this: e.g., the nature of the stimulation employed may have produced other complicating effects; or the amount of practice obtained by these animals in their normal swimming may be such that additional exercise will yield rapidly diminishing returns.

eral was markedly delayed. During the period of isolation there was a minimum of the sex behavior ordinarily displayed by monkeys at that age. At the age of 18 months, the period of isolation was discontinued, and the monkey was subsequently brought up with other members of the species. At this time, sex behavior began to appear; but in a very rudimentary form. Attempts at copulation were very crude and trial-and-error was exhibited. Sexual activity was shown indiscriminately toward males and females, as well as toward monkeys of other species, rags and other soft objects, and the experimenter's arm and hand. With continued association with other members of the species, normal sexual activity eventually developed. Other forms of behavior, such as feeding, play-activity, and grooming, were also affected by the prolonged period of isolation.

On the whole, sexual behavior is more closely dependent upon physiological factors in lower mammals, and shows an increasing modifiability and susceptibility to experiential factors in higher forms (2). For example, lower mammals such as rodents can generally copulate successfully on the first trial. This is less true among monkeys and still less among the anthropoid apes. It has been observed that some male chimpanzees have to learn to copulate, their first attempts being usually unsuccessful (40). Mention may also be made in this connection of the extensive survey conducted by Kinsey and his co-workers (26) by means of a carefully planned interview technique. This study indicated that, in the human male, sexual activity may be manifested in a wide variety of ways. The role of cultural factors in determining the occurrence of different types of sexual behavior is also suggested by the findings.

A number of experiments have been conducted on the hoarding behavior observed in many animals (31). For example, rats which are allowed access to a supply of food will carry to their home cage and store a much greater quantity of food than they can consume. Adult well-fed rats will hoard 5 to 20 pellets a day, although they eat only one or two each day. This characteristic behavior is probably associated with physiological factors, such as the sugar supply in the body. It has been shown to increase with low temperature and with food deprivation, both of which produce bodily conditions which may be conducive to the hoarding behavior. At the same time, it is interesting to note that experiential factors also seem to affect and modify this behavior. Thus some experimental data suggest that peri-

ods of food deprivation in infancy may influence adult hoarding behavior in rats (21, 22).

The experiment by Kellogg and Kellogg (25), in which a young chimpanzee was reared for a short period in a typically human environment, provides a further illustration of the role of environment in behavioral development. The chimpanzee, a female named "Gua," was isolated from its mother at the age of 7½ months and brought up in the company of the investigators' own son, then 10 months old. The association was continued for a period of nine months. The chimpanzee was not treated as a pet, but as a child, and the two subjects were given as nearly identical care as possible.

Fig. 43. Gua: A Chimpanzee Reared for Nine Months in a Typically Human Environment. (From Kellogg and Kellogg, 25, p. 226.)

Gua was clothed in the same manner as the child, and showed no difficulty in keeping on shoes, stockings, and other common articles of clothing. She slept in a bed with the usual accessories, such as sheets and blankets. Upright locomotion, not normally found in chimpanzees, was also acquired by this animal. Excellent progress was made by the ape in learning to eat with a spoon and drink out of a glass, as illustrated in Figure 43. She was able to manipulate pencil and paper to produce simple scribblings. Gua also learned to respond to oral language, and by the termination of the experimental period understood over fifty words or simple phrases, such as: "Blow the horn" (in the car); "Show me your nose"; "Do you want to go bye-bye?" "Take it out of your mouth." The degree to which it proved possible to "humanize" the behavior of this ape is indeed suggestive, especially in view of the fact that the period of residence in the human environment was of relatively short duration and did not begin at birth.

A similar experiment was subsequently undertaken by Finch.[5] A

male chimpanzee, "Fin," was removed from its mother within one day of birth and kept in a human environment until the age of 2 years and 3 months. Two human siblings were present in the environment, both older than the chimpanzee. When the ape was taken into the experimenter's home, the experimenter's daughter was four years old and the son was one year of age. In general, the findings in the case of Fin were similar to those obtained with Gua in respect to such everyday behavior as wearing clothing, eating, and sleeping. Figure 44 shows the animal at play with one of his human companions. Erect locomotion was mastered, but the structural characteristics of the ape's body made it easier for the animal to walk with the aid of his arms. Attempts to investigate responses to language met with failure, a finding whose interpretation is somewhat ambiguous. Fin's motor development was close to the norms for laboratory-reared chimpanzees, but his performance on a number of typical laboratory learning problems was poorer than that of most laboratory apes. It is apparent that at least one reason for this deficiency is to be found in Fin's poorer motivation in such situations. His attitude toward these problems was usually playful; he showed little interest in the incentive, was easily distracted, and tried to engage the experimenter in play. Such behavior suggests that Fin's superior nutritional condition as well as his experiences with humans may have actually made him a poorer subject for laboratory experiments of this sort. Motivational factors were also cited among

Fig. 44. Fin: A Chimpanzee Reared for Two Years in a Typically Human Environment. (Reproduced from unpublished material by courtesy of Dr. Glen Finch.)

[5] The writers are indebted to Dr. Glen Finch for making available the unpublished manuscript reporting this investigation.

the possible explanations for Fin's relatively poor showing in the understanding of language.

Routine habits of eating, wearing clothing, and the like were retained well, as indicated by tests made two years after removal from the human home. A vigorously displayed emotional attachment to humans and a lack of interest in other chimpanzees also persisted upon return to the laboratory. In fact, the investigator reports that Fin's friendlier and more intimate reaction to humans was his most conspicuous difference from the laboratory-reared chimpanzees. Responses to novel and unusual stimuli also differentiated Fin from the other laboratory apes, the former showing approach and active manipulation when confronted with new objects, while the latter exhibited fear and withdrawal behavior.

THE METHOD OF CO-TWIN CONTROL

Training experiments on human infants by the method of co-twin control represent essentially the same approach to the heredity-environment question as the studies on infrahuman subjects discussed in the preceding section. In such experiments, one member of a pair of identical twins is subjected to intensive training in some activity, while the other is used as a control subject and prevented from exercising the function under investigation. In one such experiment conducted by Gesell and his co-workers (cf. 19), stair-climbing and "cube behavior" (including prehension, manipulation, and constructive play with cubes) were studied in a pair of identical female twins, 46 weeks old at the beginning of the experiment. The trained twin (T) was put through a daily 20-minute training period in both types of activity for six weeks. At the end of this period, the control twin (C), who had had no specific training in these functions, proved equal to T in cube behavior. In stair-climbing, a difference was found. Whereas T was a relatively expert climber, her sister could not reach the top of a five-tread staircase even with assistance. Two weeks later, however, still without any previous training, the control twin was able to climb to the top unassisted. At this age (53 weeks), twin C was herself given a two-week training period, at the end of which she approximated T in her climbing skill. Thus, because of the higher level of maturational development, a two-week training period at 53

weeks of age proved to be nearly as effective as a six-week training period at 46 weeks.

The same pair of identical twins was subsequently put through similar training experiments in other functions, including vocabulary training (20, 38). Beginning at the age of 84 weeks, twin T was given daily, intensive training for five weeks in naming objects, executing simple commissions, and other vocabulary-building techniques. Twin C was deprived of all opportunity to hear language during that period. At the end of the five-week period, when the twins were 89 weeks old, similar training was given to twin C for only four weeks. After this training, twin C had a vocabulary of 30 words. Twin T's vocabulary at the end of her first four weeks of training had been 23 words, although at the end of the total five-week period it rose to 35 words. The investigators emphasize the role of maturation in these results, calling attention to the fact that the twin whose training was begun at the age of 89 weeks progressed more rapidly almost every day and showed a more mature manner of responding at corresponding stages of training than the one whose training was begun at 84 weeks.

It should be noted in interpreting these results that, first, the difference in rate of learning between the two twins was slight. For example, C's total vocabulary after 27 days of training equaled that of T after 31 days of training (viz., 29 words). Secondly, in vocabulary, pronunciation, and sentence construction, the twin with five weeks of earlier training slightly surpassed the one with four weeks of later training, although this difference had largely disappeared three months later. Finally, it is obvious that a certain amount of structural development in infancy and early childhood facilitates the earliest stages in the acquisition of language. The child cannot produce combinations of sounds resembling those of adults until his auditory and vocal mechanisms permit a certain degree of sound differentiation and control. Thus "maturational" factors might operate in this purely vocal aspect of language development, while the "symbolical" or meaningful aspects of linguistic development may well depend upon learning. The development of the language function may in this one respect be analogous to the development of the pecking response of chicks, as analyzed by Cruze (cf. p. 168). It will be recalled that Cruze found the effect of maturation to be limited to only one of the specific reactions which entered into the pecking behavior.

Follow-up observations of the same pair of twins through the age

of 14 are reported by Gesell and Thompson (19). The authors give a detailed summary of the development of these twins from infancy to adolescence, covering physical characteristics, motor functions, adaptive behavior, language, and personal-social behavior. The twins showed a slight difference in developmental rate and in motility, and a somewhat larger difference in sociability. In their interpretations, the authors attribute these differences primarily to innate factors. They call attention to certain environmental differences—for example, a difference in the stepmother's attitude toward the two twins—but they argue that such differences were a result rather than a cause of the dissimilarities in the behavior of the twins themselves. Other environmental discrepancies, such as a pronounced difference in the personalities of the teachers which the two twins had from the first to the fourth grade, did not, according to the authors, create any permanent psychological difference between the twins. It is their hypothesis that differences in "physiological tempo" and "ontogenetic timing," resulting from slight initial differences in the "mechanisms of symmetric regulation," must have imposed at least subtle differences in the genetic constitution of the two twins (19, pp. 105–116).

A number of points should be noted in evaluating this hypothesis. First, it is obvious that the authors do not mean "hereditary" when they use such terms as "innate" and "genetic," since the heredity of the twins was identical. They are apparently referring to prenatal environmental influences which affected the relative rate of structural development of the two subjects Their comparison, then, is between such *prenatal* environmental differences and *postnatal* differences in training and other aspects of the psychological environment. Secondly, the evidence cited is equally consistent with the hypothesis that such postnatal differences account for the behavior discrepancies. Since differences in the postnatal environment were admittedly present, the question of what was cause and what effect is debatable. Finally, if some unknown and undeterminable prenatal factors made the twins unlike in structural development, as the authors suggest, then their study offers no special advantage over the study of ordinary siblings in the analysis of the origin of behavior differences. If both learning and structural factors varied, we obviously cannot attribute the difference in behavior unambiguously to one or the other influence.

Experiments on intensive training in infancy have also been con-

ducted by McGraw (30). A pair of male twins [6] were observed from birth to the age of 22 months. Jimmy, the twin who appeared stronger and better developed at birth, served as the control, his activity being approximately that of a normal infant during the earlier period, and possibly a little more restricted than normal later. The other twin, Johnny, was put through intensive daily training from the age of 20 days. Both twins lived at home but were in the laboratory between 9 and 5 o'clock for five days a week. The performance of the trained twin in each task was compared throughout the period of the experiment with that of the untrained control twin.

Specific exercise was found to have little or no effect upon a group of activities including simple reflexes, such as suspension-grasping, as well as crawling and creeping, erect walking, sitting, prehension, and other sensori-motor functions. Marked improvement resulted, however, from practice on a group of somewhat more complex functions such as skating, jumping, swimming, diving, ascending and descending inclines, getting off stools, and manipulating and climbing stools and boxes to reach an objective. Although a certain amount of sensory and muscular development obviously helps in the latter functions, their performance seems to depend largely on specific training. The independence of the former group of functions from practice confirms many of Gesell's findings. For the execution of these simpler functions, the presence of structures of a certain degree of development seems to be sufficient or nearly sufficient.

In recent years, a few reports have appeared which refer to extensive research projects being conducted by the method of co-twin control at the Psychological Laboratories of the Moscow Medico-Biological Institute (27, 28). In one investigation, for example, 5 pairs of identical twins between 5½ and 6 years of age received 2½ months of intensive training in block building by two different methods, designated as the "method of elements" and the "method of models," respectively. In the former method, the subjects were allowed to see the individual blocks making up the figure which they were to copy. In the latter, the sample block figure was covered with paper, thus making the individual blocks indistinguishable. One member of each pair of identical twins was taught by the former, the

[6] Originally believed to be identical, although subsequent physical development made this designation unlikely.

other by the latter method. Those taught by the "model" method did more poorly on the test itself, but excelled on other block-building tests as well as on a number of other visual-perceptual and spatial tests. These differences persisted in tests administered 10 months after the cessation of training. Such results suggest the possible role of *work methods* in the development of individual differences in performance, a factor whose importance has been pointed out by other writers (cf., e.g., 1, 33).

EXPERIMENTAL RESTRICTION OF TRAINING IN HUMAN INFANTS

With certain limitations, restriction of training—the principal technique employed in the animal studies reported in the first section of the present chapter—has been applied in a few studies with human infants. This procedure is illustrated in an experiment by Dennis (12). Two female infants, who happened to be fraternal twins although this was irrelevant to the present experiment, were reared under controlled conditions in the experimenter's home from one to 14 months of age. During the first 7 months of this period, stimulation and activity were rigidly restricted. Opportunities for standing and sitting were eliminated, and opportunities to grasp objects were highly minimized. The nursery was bare of all but essential furnishings. The experimenters had no social contact with the children except for physical care and for a few tests made during this period. They did not smile, frown, speak to, or play with the subjects. The two infants were separated from each other by means of an opaque screen.

The subsequent behavior development of these two children was compared with norms established on infants brought up under normal, unrestricted conditions. The age at which each of the two experimental children first performed each of a number of specific activities was charted in reference to the average and range of ages at which such activities appeared in the "normal" groups. Functions which normally appear during the first seven months showed no appreciable retardation in the experimental subjects. Among these were such simple functions as fixating objects; starting, turning the head, or crying at a sound; grasping objects; watching or playing with own hands; and bringing hand or object to mouth. Evidently functions such as these

appear, regardless of exercise, when the necessary structural development has occurred. Responses which normally occur beyond the seventh month did, on the whole, show significant retardation, the age at which they appeared in the experimental subjects often falling beyond the range of the comparison subjects. The investigator reports that these responses were quickly established when the opportunity for practice was provided, and concludes that their absence had resulted from lack of self-directed practice, rather than from lack of instruction or of socially administered rewards.

CULTURAL DIFFERENCES IN INFANT-REARING PRACTICES

A sort of natural "experiment" similar to that described in the preceding section is provided by infant-cradling practices prevalent among certain cultures. It has been the custom in Albania, for example, as in a number of its neighboring countries, to bandage the children tightly to their cradles during the first year of life, so that they cannot move their arms or legs. The cradle is kept in a darkened room, and the child has no contact with toys or other objects. The infant is unswathed and bathed once a day, and sometimes less often.

Tests administered to 10 such infants between the ages of 4 months and one year (cf. 11) showed considerable behavorial retardation. Few reacted spontaneously when given the opportunity to do so. Coordination was decidedly poor, and grasping movements occurred much later than normal. Only one out of the 10 infants was able to crawl before the age of one year, although all could sit up without support. Social reactions, on the other hand, were found to be advanced, a finding which was attributed to the presence of large families and to the fact that persons were the only familiar type of stimulation in the child's environment. Children over one year of age were reported to be normal in social reactions, learning ability, and "mental productivity," but retarded in coordination and expression. Their attitude toward new objects is described as interested and willing, though shy and clumsy, the children frequently depending upon adults for help in novel situations.

Restrictive infant-rearing practices are also to be found among certain American Indian tribes, such as the Navajo and the Hopi.

Among the Hopi, the newborn child is bundled tightly in a blanket and tied securely to a stiff board. In such a position, the infant cannot move his arms or legs, or even turn his body. For the first three months he is kept in these wrappings, except for about one hour each day, when he is cleaned and bathed. Dennis (13) reports that despite this extreme restriction of movement, when Hopi children are released they show the same sitting, creeping, and walking behavior—and in the same sequence—as white American children. During the short daily periods when they are freed of their wrappings, moreover, they assume the usual flexed position, reach for objects and carry them to the mouth, reach for their toes and put them into the mouth, and exhibit other characteristic motor behavior of an unrestricted infant. It is also interesting to note that no significant difference was found in this study between the average age of walking of Hopi infants cradled in the traditional manner and other Hopi children who had been cradled in the manner of white American children. In a group of 63 children reared in the Hopi manner, the average age of walking was 14.95 months; for 42 Hopi children reared without binding, the average age of walking was 15.05. This difference is not statistically significant.

The results of these two studies are not in complete agreement, the former suggesting somewhat greater disruption of functions through lack of exercise than the latter. The specific results probably depend upon a number of factors, such as the nature and degree of the restriction and the age at which it is discontinued. The particular functions observed also undoubtedly differ in their dependence upon practice. Additional comparative data are obviously needed before we can draw any definite conclusions. In general, however, such studies suggest that certain simple, reflex functions of early infancy depend almost wholly upon structural development, while many others require a relatively brief period of exercise for their successful performance.[7]

[7] Two cases have been reported of American children who, because of wilful neglect on the part of parents had been reared for several years under much more extreme conditions of restricted activity. The interpretation of such cases is difficult, however, because of low mentality of parents and because of the very poor physical condition of the children resulting from neglect (cf. 35, pp. 249-250). A third case has been described of an infant, neglected medically and psychologically, who had a Kuhlmann-Binet IQ of 29 at the age of 19 months. The IQ rose steadily following institutional care and training, and seemed to have reached a stable level at 97 by the age of 6 (29).

CASE REPORTS OF "FERAL MAN"

A much more extreme type of "natural experiment" is furnished by the cases of children found to have been living in isolation or in exclusive contact with lower animals. Such "wild children" have been described since early historic times. In 1758, Linnaeus included them in his classification of the human species, under the designation of "feral man." An extensive survey of recorded cases has recently been prepared by Zingg (41, 42). Over forty cases are described, although in a number of them the available information is quite meager or the isolation was only partial. These wild children include a few who had apparently been abandoned or had wandered off and survived in the wild largely through their own efforts, as well as a number who seem to have been reared by such animals as the wolf, bear, goat, pig, sheep, cattle, and leopard. Children who have been confined in isolation from human contacts and have been living under conditions barely sufficient for survival are also included in this category.

These cases of wild children have been of special interest to psychologists because of the possible light they may throw upon the question of how far normal human behavior develops in the absence of normal human stimulation. In his summary of the recorded cases, Zingg (41) concludes that such wild children were, without exception, mute and quadrupedal. No vocalization resembling human speech developed under these circumstances, and the characteristically human, erect locomotion was not found. All had developed some form of locomotion on hands and feet or on hands and knees, and their physical structure had often become modified (by the appearance of calloused pads, etc.) to permit rapid and efficient quadrupedal locomotion.

Characteristic sensory modifications are also reported, the senses of smell, hearing, and sight—especially night vision—often showing an animal-like keenness. Eating habits are markedly unlike those typical of the human. Raw meat is the common diet among children reared by carnivorous animals; wild-living children are described as subsisting largely on bark, roots, grass, herbs, and leaves. One "wild girl" in France had become very adept at swimming for fish and frogs, which constituted her principal food. The pattern of eating behavior is also similar to that of lower animals, including the smelling of food before eating, lowering the mouth to the food, sharpening

teeth on bones, and the like. There is no evidence of any tendency to cover the body or to devise "clothing" of any sort. Such children seem to have been relatively insensitive to heat and cold and to have developed no "sense of shame" from nakedness. No crying, tears, or laughter was observed, although other expressions of violent anger or impatience are reported. Expressions of sex interest and activity were either completely absent or present only in the form of diffuse, general, undirected activity. No "consciousness of kind" or gregariousness was evidenced, the children shunning humans and often showing preference for the company of lower animals.

These reports of wild children have been seriously questioned by some psychologists, such as Dennis (14). It is undoubtedly true, as the records themselves show, that in several of the cases summarized by Zingg the association with animals either began after the child had reached an advanced age in human contact, or such association was only partial, the child still remaining in some contact with human adults. It is also true that the data on some of the cases, especially the earlier ones, are so meager and so subject to the inaccuracies and bias of the original observers as to be of doubtful authenticity.

Some writers (cf., e.g., 14) would go further, however, and propose an alternative explanation for *all* cases of wild children. They maintain that such children may have been feebleminded to begin with, which would account for their abandonment in certain cultures. Their lack of typical human behavior, such as language and erect locomotion, as well as their other "animal-like" characteristics, are then attributed to their original mental defect. The usual counter-argument is to ask: "How, then, could such a feebleminded child have managed to survive in an environment which would tax the ingenuity of even a normal adult?" In answer to this, Dennis (14) has proposed the possibility that the wild children may actually have been abandoned only a short time—perhaps only a few days—before they were found, and that their behavior deficiency was incorrectly interpreted as a sign of prolonged isolation.

In a reply to Dennis' critique, Zingg (42) calls attention to the structural changes following prolonged four-footed locomotion, as well as to the degree of proficiency attained in such locomotion, conditions unlikely to have developed if the child had been living in human society until a short time prior to his discovery. The marked progress in learning human ways made by a number of such children

is also mentioned as evidence against the hypothesis of initial feeble-mindedness. To be sure, one rarely finds the performance of such subjects after training to equal or even approximate that of normal children of the same age. This is hardly to be expected, however, not only because of the long period of isolation during which opportunities for acquiring normal human behavior were lacking, but also because of the interference or "negative transfer" from other modes of behavior which have been acquired in the wild and which must be "unlearned" before progress can be made. By way of contrast, Zingg calls attention to one case of a "wolf-boy" of India who was apparently a "true idiot." This boy showed virtually no progress subsequent to his capture, although he lived well into adulthood. The comparatively greater progress made by the other wild children suggests that they may have been structurally normal and rendered deficient only by their early stimulational deprivation.

Zingg also cites the reports of reliable eye-witnesses indicating that at least two wild children (the wolf-children of Midnapore, to be reported below) had actually been living with wolves for some time prior to their capture. This is emphasized by Zingg in reply to Dennis' argument that no direct evidence is available to show that human children have in fact been reared by animals. Dennis suggests that when children are captured in the company of animals, they may simply have been accidentally brought together by their common efforts to hide from a pursuer. Dennis stresses the importance of this point for the interpretation of the behavior of wild children. He points out that if a child is abandoned before the age of about three, he cannot possibly survive in the wild unless "adopted" and cared for by an animal. On the other hand, if the child was over three at the beginning of his wild existence, then he would already have acquired at least the rudiments of human speech, locomotion, and similar functions, unless he were congenitally defective. It thus becomes of crucial significance to establish the possibility that, at least during their earliest, helpless years, such children could have been nurtured by animals.

The objections raised by Dennis (14) and others must be carefully weighed in interpreting any report on allegedly wild children. Such objections may well hold for a considerable number of the cases cited by Zingg. On the other hand, the available evidence strongly suggests that at least three or four cases are well-authenticated, genuine in-

stances of prolonged isolation from human contact. One of the most intensively studied cases is that of Victor, the Wild Boy of Aveyron (23). In September, 1799, three sportsmen came upon a boy of 11 or 12 in a French forest. The boy was completely naked, unkempt, scarred, unable to talk, and seemed to have been leading a wild, animal-like existence. He was seized by the men as he was climbing a tree to escape their pursuit, and was subsequently brought to civilization, where he finally came under the guidance and observation of the French physician Itard. The very illuminating account which Itard published on his own findings has immortalized the Wild Boy of Aveyron.

When found, the boy seems to have been deficient in all forms of behavioral development, including sensory, motor, intellectual, and emotional. This is clearly brought out in the following description given by Itard (23, pp. 5-8):

> His eyes were unsteady, expressionless, wandering vaguely from one object to another without resting on anybody; they were so little experienced in other ways and so little trained by the sense of touch, that they never distinguished an object in relief from one in a picture. His organ of hearing was equally insensible to the loudest noises and to the most touching music. His voice was reduced to a state of complete muteness and only a uniform guttural sound escaped him. His sense of smell was so uncultivated that he was equally indifferent to the odor of perfumes and to the fetid exhalation of the dirt with which his bed was filled. Finally, the organ of touch was restricted to the mechanical function of grasping objects. Proceeding then to the state of the intellectual functions of this child, the author of the report presented him to us as being quite incapable of attention (except for the objects of his needs) and consequently of all those operations of the mind which attention involves. He was destitute of memory, of judgment, of aptitude for imitation, and was so limited in his ideas, even those relative to his immediate needs, that he had never yet succeeded in opening a door or climbing upon a chair to get the food that had been raised out of reach of his hand. In short, he was destitute of all means of communication and attached neither expression nor intention to his gestures or to the movements of his body. He passed rapidly and without any apparent motive from apathetic melancholy to the most immoderate peals of laughter. . . . His locomotion was extraordinary, literally heavy after he wore shoes, but always remarkable because of his difficulty in adjusting himself to our sober and measured gait, and because of his constant tendency to trot and to gallop. He had an obstinate habit of smelling at anything that was given to him, even the things which we con-

sider void of smell; his mastication was equally astonishing, executed as it was solely by the sudden action of the incisors, which because of its similarity to that of certain rodents, was a sufficient indication that our savage, like these animals, most commonly lived on vegetable products.

It is interesting to note that the sensory deficiency of this boy seems to have been quite specific and in many instances directly traceable to his mode of life. Thus Itard observed that "the sound of a cracking walnut or other favorite eatable never failed to make him turn around . . . nevertheless, this same organ showed itself insensible to the loudest noises and the explosion of firearms" (23, p. 15). Sexual development showed the same general undifferentiated type of response observed in the case of animals reared in isolation. Following the onset of puberty, periods of vague restlessness and discomfort as well as occasional fits of sadness or anger were noted, without, however, the development of specific, normal sexual activity.

After five years of ingenious, painstaking, and methodical training, Itard abandoned the task because he had failed to bring the boy up to a normal level of performance. It is significant, however, to note the degree of improvement which was effected during this period. Itard himself writes: "But if one limits oneself to the two terms of comparison offered by the past and present states of young Victor, one is astonished at the immense space which separates them; and one can question whether Victor is not more unlike the *Wild Boy of Aveyron* arriving at Paris, than he is unlike other individuals of his age and species" (23, p. 53). Besides learning many routine activities of a civilized community, including eating habits, dressing, personal care, and the proper use of common articles of furniture, Victor showed considerable progress in the identification and discrimination of objects, the formation of simple abstract concepts, and other intellectual tasks set by his tutor. Although unable to articulate sounds, he succeeded in learning to communicate through written language, being able through this medium "to express his wants, to solicit the means to satisfy them and to grasp by the same method of expression the needs or the will of others" (23, p. 84). Evidence of considerable development in emotional responses and in social and "moral" attitudes is also cited by Itard.

The more recently discovered "wolf children" of Midnapore (18, 24, 36, 37, 41) represent another well-authenticated case. In 1921 two girls, one approximately two to four years of age and the other

eight or nine, were found living in a cave with wolves in a sparsely settled region of India. They were taken into a local orphanage, where attempts were made to train them. A detailed diary of the girls' activities, kept by the rector of the orphanage, is now available in published form (35), together with analyses and comments by several psychologists, a sociologist, a geneticist, and an anthropologist. It proved difficult to keep the girls in good health, particularly because the readjustment to a normal human diet led to physical debility and severe skin reactions. The younger girl, "Amala," died within a year; the elder, "Kamala," lived for about eight years after her discovery. Most of the observations on record are thus necessarily based on the elder child.

Kamala, like her younger companion, showed a strong preference for raw meat, and was fond of pouncing upon any freshly killed animal which she found. Displaying a keen, animal-like sense of smell, she was able to detect the odor of meat from a great distance. Hearing was also very acute. Her eyes are described as having a peculiar glare, like the eyes of dogs or cats in the dark. It seemed that Kamala could see better at night than in the daytime, and she seldom slept after midnight. It has been suggested (18) that the vitamin intake in her diet may have favored chemical changes in the eye which improved vision in dim light. Eating and drinking were accomplished by lowering the mouth into the plate. In general, her mouth, rather than her hands, served as a prehensile organ. Eventually she was taught to use her hands in eating.

As in the case of other feral children, locomotion was quadrupedal. Kamala walked on hands and knees for slow locomotion, and on hands and feet for running. She was able to run so fast by this method that it proved difficult to overtake her. Thick callosities had developed at the knees, elbows, soles, and palms, undoubtedly as a result of such locomotion. It was not until six years later that she finally adopted erect walking, although even at that time she would revert to the former four-footed technique when running. Her only vocalization at the time of discovery consisted of a cry or howl which bore a certain resemblance to the typical wolf cry. With prolonged training, she was finally able to say about forty-five words and to form simple sentences of two or three words each.

Mention should also be made of the celebrated and mysterious case of Kaspar Hauser (cf. 35, pp. 277–365), about whom so much

has been written. Some accounts suggest that this boy was an heir to a princely house and was put out of the way by political enemies. He was apparently confined from early childhood in a dark cell, not large enough for him to stand upright. No clothing or cover was furnished except a shirt and trousers. When he awoke, he was accustomed to find bread and water, but he never saw the person who brought them and he had no knowledge of the existence of other living creatures besides himself. He was released in 1828, when about 17 years of age. At this time he was first discovered wandering aimlessly about the streets of Nuremberg. He could not talk, but repeatedly uttered certain phrases meaninglessly. He is reputed to have had a remarkable sense of smell and a surprising ability to see in the dark. His walking resembled the first efforts of a child. After various vicissitudes, his instruction was undertaken by a skillful and painstaking teacher. Under the latter's tutelage, Kaspar Hauser made rapid progress and soon learned to speak. By this means he was able to communicate what he recalled of his life in the cell as well as his experiences during his period of instruction. Unlike other cases of children brought into contact with civilization relatively late in life, Kaspar Hauser profited sufficiently from his education to reach and even possibly surpass normal achievement.

To psychologists, probably one of the most interesting observations made in the course of Kaspar Hauser's training is one pertaining to the development of space perception.[8] Shortly after his release, Kaspar Hauser was asked by his tutor to look through a window which gave a view of a pleasant landscape. The boy is reported to have withdrawn in horror, saying, "Ugly, ugly!" At a later date, when questioned about this incident, Kaspar Hauser explained it as follows:

> Yes, indeed, what I then saw was very ugly For when I looked at the window it always appeared to me as if a window-shutter had been placed close before my eyes, upon which a wall-painter had splattered the contents of his different brushes, filled with white, blue, green, yellow, and red paint, all mingled together Single things, as I now see things, I could not at that time recognize and distinguish from each other. This was shocking to look at; and besides, it made me feel anxious and uneasy; because it

[8] This incident is reported in the contemporary account written by Paul J Anselm, Ritter von Feuerbach, Kaspar Hauser's protector and friend. This account, which is considered to be the most authentic source of information on the case, was translated into English in 1833 by H G Linberg and is reprinted in full in Zingg's book (35 pp 277 365).

appeared to me as if my window had been closed up with this parti-coloured shutter, in order to prevent me from looking out into the open air. That what I then saw were fields, hills, and houses; that many things which at that time appeared to me much larger, were, in fact, much smaller, while many other things that appeared smaller were, in reality, larger than other things, is a fact of which I was afterwards convinced by the experience gained during my walks; at length I no longer saw anything more of the shutter (35, p. 323).

In a further quotation from the same account, we are told:

To other questions, he replied, that in the beginning he could not distinguish between what was really round or triangular, and what was only painted as round or triangular. The men and horses represented on sheets of pictures, appeared to him precisely as the men and horses that were carved in wood; the first as round as the latter, or these as flat as those. But he said, that, in the packing and unpacking of his things, he had soon felt a difference; and that afterwards, it had seldom happened to him to mistake the one for the other (35, p. 323).

These examples of "wild children" illustrate the close dependence of human development upon the environment in which the subject is reared and the type of stimulation to which he is exposed. If a child is deprived of normal human contacts, his behavior will come to resemble in many ways that of a low-grade idiot. Such a condition has, in fact, been regarded as a sort of environmental feeblemindedness and has been given the name of *isolation amentia* (cf. 39, pp. 292–297). When a child is brought up in contact with animals, striking similarity to the behavior of those animals is exhibited, and such behavior proves difficult to eradicate once it has become firmly established. Subsequent educational efforts are inadequate to undo the effects of early nurture. Rousseau's dream of the "noble savage" whose inner nature is allowed to develop, free and unhampered by human interference, proves to be a vain chimera. The situation has been aptly summarized by Stratton (37, p. 597):

Lack of association with adults during a certain critical period of early childhood, it seems likely, produces in some or all normal children marks like those of congenital defect. The evidence seems against the romantic view that a civilized community is a chief obstacle to the development of personality. On the contrary, the higher forms of personality become possible only in and through such a community. By our biological endowment

CHAPTER 7

The Effects of Practice

AN UNDERSTANDING OF THE DEGREE to which scores on a psychological test can be raised by practice or by coaching is obviously essential for the proper use and interpretation of such a test. Apart from this purely technical consideration, however, the investigation of such practice effects is relevant to the role of stimulational differences in behavior development. This type of investigation has been treated by some psychologists as another approach to the pervasive question of heredity and environment.[1]

Theoretically, the interpretation of the effects of practice or coaching upon mental test performance presents several possibilities. Thus, some have argued that if psychological test performance should prove relatively impervious to improvement through practice or coaching, then the scores on these tests may be regarded as indices of "native ability" or "potentiality." If on the other hand, the tests do prove susceptible to such influences, then it might be argued either that these particular tests are unsuitable as measures of "native capacity," or that the concept of "native capacity" should be redefined or discarded. However inconsistent some of these interpretations may be with the nature of heredity and environment, as discussed in Chapter 3, it is necessary to keep them clearly before us, since all such interpretations have been expressed in the highly controversial literature of this field. Failure to recognize the shades of difference among these views has added to the confusion in discussions of heredity and environment

A further question which has been raised regarding the role of practice or coaching relates to the *permanence* of the effects. As con-

[1] Cf. the seventh method listed in Chapter 4.

ceived by some writers, this question pertains to whether the "underlying course of development" may be altered by environmental conditions or whether such conditions exert only a "superficial" and "transitory" effect Still another way in which the results of such experiments have been analyzed is in terms of the effect of practice upon the *extent of individual differences*. Do individuals become more alike, or do they become more unlike each other when subjected to a uniform period of practice? The implications of the answers to this question have also been disputed at length, and will be considered in the closing section of the present chapter

THE EFFECTS OF PRACTICE UPON PERFORMANCE LEVEL

All the investigations to be considered in the present section were conducted on children of school age Gates (7) studied the effect of continued practice upon *memory span for digits* Two groups of school children selected so as to be equivalent in age, number of boys and girls, Stanford-Binet IQ, school grade, teachers' estimates of scholastic maturity, and scores on several memory tests, were given an initial test in digit span The children in the Practice group were then put through individual practice in recalling digits on each of 78 days extending over a period of five months At the end of this period, both Practice and Control groups were given a final test The average scores on initial and final tests are reproduced in Table 4

TABLE 4 *The Effects of Practice upon Memory Span for Digits*

([illegible])

Group	*Initial Test*	*Final Test*	*After a Lapse of 4 Months*	*After Common Practice for [illegible] Days*
Practice	4 33	6 40	4 73	5 73
Control	4 33	5 06	4 83	5 92

Both groups show improvement, but the Practice group is clearly ahead, manifesting a gain which normally requires a six-year period, according to the Stanford-Binet norms for this function Four and one-half months after the final test, both groups were again tested, by a

different examiner. This time the Practice and Control groups were approximately equal. Finally, the two groups were subjected to 22 days of practice, at the end of which both showed improvement and in approximately equal degree. It was also found that the training in digit span had no effect on performance in other types of mental tests. From these findings, Gates concludes that training is highly specific, consisting in the acquisition of special skills and techniques, and that it does not alter the growth of the "underlying mental functions."

It is unfortunate that, in most studies on experimentally administered practice, all effects not directly resulting from such practice are attributed indiscriminately to growth or maturational phenomena, the influence of the vast amount of other training which the child is receiving in the course of everyday life being disregarded. Thus in Gates' experiment, the improvement of the Control group could just as well be attributed to intervening experiences as to growth. The data themselves, to be sure, do not permit a choice between the two explanations. That experience may have been the important factor is, however, suggested by the fact that both groups drop to an equal level when retested later by another examiner. The drop may have resulted from the time of year at which the tests were administered, or from other factors incidental to the school situation. The closeness with which the child attends to the material and the effort he puts forth to concentrate on the task of memorizing are very important factors in determining his span; and it seems entirely plausible that such factors should be influenced both by the attitude of the particular examiners and by the sum total of school experiences which the child has had. It is noteworthy that the 4½-month period preceding the drop in score included the summer vacation, which is definitely an environmental and not a maturational incident.

The marked susceptibility of a function like memory span to training, which this experiment demonstrated, seems in itself to minimize maturational factors. To assume the existence of some underlying hypothetical capacity of memorizing which remains unaltered while performance on a memory span test rises and fails seems totally unwarranted by the facts and certainly does not clarify the problem.

The effect of repetition upon *intelligence test scores* has also been investigated. Rises in score have been regularly reported when the *identical test* is repeated within periods ranging from a few days to a

year. In a survey of 614 school children, for example, Adkins (1) found that children taking three common group tests for the second or third time at intervals of a year obtained higher scores than children of corresponding school grade who were taking the tests for the first time. Similarly, in the Harvard Growth Study (6), the median IQ in successive years rose from 102 to 113 when the same group intelligence test was repeated,[2] but dropped again to 104 when another group test was substituted. In the report of this study, it is pointed out that the meaning of an IQ on repeated tests may change considerably. Thus, for example, an IQ of 100 may correspond to the 47th percentile on the first testing and to the 17th percentile on a subsequent testing (6, p. 134). In other words, an IQ which, if obtained on a first test, would indicate approximately average ability, on a retest might signify ability in the lowest quarter of the distribution. In an analysis of the Stanford-Binet IQ's of children in the same study (cf. 5), retests administered within six months showed an average gain of four to five IQ points.

Gains in score have also been found upon the administration of *parallel forms* of the same test, although such gains tend in general to be smaller. Terman and Merrill (21) report an average increase of approximately 2.5 IQ points when Form L of the revised Stanford-Binet was followed by Form M, or vice versa, within a few days. In the Minnesota Preschool Scale, it is suggested that 3 IQ points be deducted as a correction for practice effect when alternate forms are administered within a few weeks (cf. 8).

E. L. Thorndike (23) gave alternate forms of a group intelligence test to several groups of high school, college, and graduate students. The two forms were administered in immediate succession and were preceded by a 10-minute fore-exercise on similar items. The forms were used in reverse order in different groups so as to cancel any existing differences in the difficulty of the parallel forms. The average gain in score on the second form was approximately 8 points for the various groups tested. In an earlier study in which the fore exercise had been omitted, the average gain had been slightly over 12 points. In a further investigation with the same test, 15 equivalent forms were

[2] The authors report that the gains lessened in later trials, because the brighter subjects were reaching the upper limits of performance artificially set by the test ceiling. Thus the gains would presumably have been larger if tests with higher ceilings had been employed.

administered, one each day, to 20 gifted and 19 normal children, all approximately 11 years old and attending the same school. The mean of the gifted group rose from 87.5 to 111. The largest gain occurred on the second trial, and the highest mean score, 115, was reached on the 10th trial. Subsequent trials showed only minor fluctuations. In the normal group, the mean rose from 51 to 86.5, the largest gain again occurring on the second trial. Subsequent trials showed smaller fluctuations, in either direction, with the highest mean score falling on the 13th trial.

Some evidence is available suggesting that a slight improvement also occurs upon successive retests with *different intelligence tests*. Rodger (18), for example, gave six group intelligence tests to 76 children, aged 11 and 12, with an interval of two weeks between each successive test. The children's average IQ rose from 101.9 on the first test to 109.8 on the sixth.[3] Although this practice effect was found at all ability levels, the brighter children, in general, showed larger gains. For example, the 95th percentile rose from 123.5 to 138.5, whereas the 5th percentile rose only from 81.2 to 83.0. Another investigator (6), however, found no such practice effect from one intelligence test to another, the improvement being specific to the particular test. Undoubtedly the "spread" of the practice effect will differ with such factors as the degree of similarity of the tests, and the age, education, and previous "test-wiseness" of the subjects.

A *qualitative analysis* of the changes occurring when tests are repeated was undertaken by Greene (10, 11). Groups of from 19 to 235 college sophomores were given four trials of each of 14 tests at intervals of one day, not all tests being employed with any one group. A wide range of fairly specific functions was covered by the tests, which included the Seashore musical discrimination tests, tapping, aiming, pencil mazes, digit span, feature comparison, speed of reading, equation completion, vocabulary, Kohs Block Design, Stenquist Mechanical Assembly, and Minnesota Spatial Relations. An analysis of the test scores, supplemented by observations of performance and introspective reports, led the author to conclude that the qualitative changes in procedure correspond closely to the differing amounts of improvement on the various tests. Those tests which showed little or no

[3] These scores were adjusted for possible differences in norms resulting from the fact that different tests had been standardized on different populations

improvement with repetition depended upon processes which change little with practice; such tests were performed in like manner on initial and subsequent trials. Tests showing large practice effects, on the other hand, seemed to depend upon different processes or work methods when repeated, some processes being eliminated and new ones introduced.

Tests involving primarily speed of ballistic movement (e.g., tapping) or liminal discrimination (e.g., the Seashore tests) showed the smallest practice effect, the improvement ranging from 0 to 5 per cent of the initial scores. Tests depending upon precision of movement (e.g., aiming) or upon specific preliminary information (e.g., vocabulary) showed increases of 6 to 25 per cent. At the other extreme, increases of 76 to 200 per cent were found in such tests as the Kohs Block Design and the pencil mazes, in which a generalized rule or principle could be learned during the performance of the test. Increases of more than 300 per cent occurred in those tests in which a solution or partial solution could be recalled and applied directly in subsequent trials, as in the Stenquist Assembly Tests.[1]

It is apparent that, at least in certain types of tests, repetition may produce a change in the nature of the test. Thus a test which on its first administration measures arithmetic reasoning or mechanical aptitude may, upon repetition, become primarily a test of memory and speed. On later trials the subject need only recall and execute the finished solutions which he worked out during the initial trial. It might be argued that this is simply a problem of test administration and has no bearing upon the development of behavior. Obviously the norms, validity and reliability determined for the initial trial of such a test would be inapplicable to the scores obtained upon repeated testing. Repetition of such a test might be said to "spoil" it as a diagnostic instrument. According to this view, it is only the measuring instrument which is affected and not the behavior which it is designed to measure.

In evaluating such an interpretation, it should be noted, first, that any *dichotomy* between "test behavior" and "underlying behavior functions" is misleading and inconsistent with the logic of test construction. Every psychological test necessarily samples behavior func-

[1] It should be noted, of course, that these differences in per cent gain may result in part from differences in the location of the arbitrary zeros of the various tests.

tions. Any influence which affects test performance, therefore, may also affect performance outside of the test situation. If repetition of a test leads to marked improvement, then repetition of similar activities in everyday life will probably lead to marked improvement in the performance of such activities.

Secondly, if repetition of a test alters the nature of the behavior being sampled, because *different work methods* are employed in taking the test before and after practice, then the role of work methods ought also to be considered in comparing the *initial performance of different individuals*. For example, individuals whose previous experience includes the solution of many arithmetic problems dealing with amount of money spent and saved out of weekly earnings, or number of pencils which can be bought for a given amount of money, will rely more heavily on memory and routine solutions, and less heavily upon reasoning, in taking a test which consists of such problems. The reverse will be true of individuals without such previous experience. This is even more apparent in such tests as mechanical assembly. On a test of this sort, the initial performance of a person who has frequently taken apart and put together bells, clocks, latches, and other mechanical gadgets may be more nearly comparable to the third trial performance of a mechanically inexperienced individual than to the latter's first-trial performance.

In other words, when a test proves to be markedly susceptible to practice effect, the behavior which it measures is probably susceptible to practice effect in everyday life, to a corresponding degree. Individual differences in such behavior may therefore result largely from such differences in previous experience. This does not mean, of course, that repetition of the test—or any other factor which raises test performance—will in itself improve the behavior area which is being sampled. Thus it would be absurd to expect that the repetition of a particular mechanical aptitude test, which raises the subject's score from the 40th to the 70th percentile, has increased his general mechanical aptitude by that amount. The subject's mechanical aptitude has been raised only for the small sample of tasks included in the test and any similar tasks to which he can apply the specific procedures or principles he has thus learned. But such a rise in score does suggest that similar practice in daily life might raise performance in the broader area which the test is sampling.

THE INFLUENCE OF COACHING

A number of investigations have been conducted to determine the effect of specific coaching upon intelligence test scores. The 1916 revision of the Stanford-Binet has probably been more thoroughly explored in this regard than any other test. In three studies (4) carried out under the direction of Terman at Stanford University, children were given instruction and practice for several weeks on material either identical or similar to some of the tests in the Stanford-Binet scale. The groups were small, varying from 10 to 26, but in each study the trained group was carefully matched with a control group by "pairing" the subjects. All experiments clearly demonstrated the possibility of teaching a child to perform tests which he was formerly incapable of doing because of age or mental level. The influence of this improvement upon the IQ obtained on the whole scale differed in the three studies, being most evident, as would be expected, in that study [5] in which the trained functions overlapped with the largest number of Stanford-Binet tests. In this study, furthermore, retests after a six-week period, during which neither group had received any training, showed the practice group to have retained its advantage over the control group.

A more extensive investigation on the effects of coaching is reported by Greene (9, 12). Three groups of children were given the Stanford-Binet. The subjects in one of these groups were then coached on the specific tests in which they had failed. A second group was coached on material similar but not identical to that in the Stanford-Binet. *No child in either group was coached for over two hours altogether*, although the training was distributed over a period of two weeks. The third group served as a control, receiving no special training in the test material. All groups were retested at intervals of three weeks, three months, one year, and three years after the initial tests. The average IQ's of each group on the initial test and on each of the four retests are given in Table 5. The results obtained in two schools, A and Y, have been kept separate since a slightly different method of coaching was employed in each. A total of 153 second grade school children served as subjects during the first year of the study, but only 83 could be reached in the three-year retest. The data reproduced below are based only on the subjects tested over the entire three-year period.

[5] I.e., the study by Casey (4, pp. 431-433).

TABLE 5 *The Effects of Coaching*

(From Greene, 12, p. 425)

Test	*Average IQ: School A* Control (N = 9)	Coached (N = 18)	Similar (N = 17)	*Average IQ: School Y* Control (N = 17)	Coached (N = 11)	Similar (N = 16)
I. Initial	82.33	84.22	101.35	98.05	98.55	101.06
II. 3 weeks	88.22	107.94	109.47	100.18	133.09	107.81
III. 3 mos.	87.78	103.17	113.41	97.76	114.55	104.31
IV. 1 year	86.56	94.28	106.76	100.40	113.73	106.88
V. 3 years	85.44	88.67	106.71	96.18	102.82	98.75

It will be noted that whereas the control groups manifest only irregular fluctuations from time to time, the coached groups in both schools show marked improvement on the second test, which followed shortly after the coaching period. This improvement is retained on successive retests, although in constantly decreasing amount. The gradual drop in IQ observed in the coached groups may be attributable partly to forgetting of the coached material and partly to the fact that, as the children grew older, they were tested to an increasing extent at higher age levels *in which they had not been coached.*

That the latter is probably the major factor is demonstrated by a comparison of the coached groups in the two schools. In school A, the children were coached more intensively on fewer tests; in school Y, they were coached on two additional higher levels. Thus the effects of coaching in school Y should not be "outgrown" as readily as in A. The average IQ's do in fact show larger and more lasting effects of coaching in school Y. The groups trained on similar material also show an immediate improvement, which gradually disappears on successive retests. As would be expected, the gains in these groups are much smaller throughout than in the groups which had been directly coached.

All these studies indicate the great extent to which mental test performance may be influenced by training. Such findings suggest vast possibilities regarding the part played by the incidental and often accidental training of everyday life. That the effects of a brief period of training are not permanent seems to be quite beside the point. *When training is discontinued,* we should naturally expect the improvement to fall off because of forgetting. If, furthermore, children are tested in different functions at successive ages, as they are to a large extent in

the Stanford-Binet, the effects of training will not be manifested over a long period. It is futile to expect that a brief period of highly specific instruction or practice should raise the "general mental level" of the child, especially since such a mental level is itself a manifold of widely diverse and loosely interrelated functions. Training does have a very real effect, however, upon the individual's performance on specific mental tests. And this is of prime importance since all our observations regarding the subject's psychological make-up are ultimately derived from such concrete behavior.

THE PROBLEM OF PRACTICE AND INDIVIDUAL DIFFERENCES

Since it has been demonstrated that training can bring about a pronounced change in mental test performance, a further question may be raised regarding the differential effects of such training upon individual subjects. Will the initially better individuals benefit more than the initially poorer? Will subjects tend to maintain the same relative standing in the course of training? Do individual differences increase or decrease with practice? If these questions are still unanswered, it is not for dearth of data, for they have been repeatedly investigated with a wide variety of materials, methods, and subjects.[6] The entire problem is so beset with technical difficulties, however, as to have even been declared insoluble by some. The crux of the matter is that entirely opposite conclusions can be drawn if the results are expressed in different forms, a fact which has cast an aura of artificiality over all the data.

In the present section, we shall examine briefly some of the major issues involved in the problem of practice and variability. These must be considered before any attempt can be made to interpret particular findings. The data are meaningless unless evaluated in terms of the specific questions which we wish to answer and the methodology necessitated by such questions. This section may seem somewhat of a technical digression, but it cannot be eliminated from any analysis of the effects of practice upon individual differences. Attempts to present only a simplified summary of results have proved exceedingly mis-

[6] For summaries of the relevant literature, the reader is referred to Kincaid (15), Peterson and Barlow (16), Reed (17), Anastasi (2), Burns (3), and Yoshioka and Jones (25).

leading, since the reviewer in such cases must either arbitrarily omit many of the data or offer conflicting conclusions with no possibility of reconciling them.

Many of the difficulties met in this problem are inherent in any comparison of variability, either from trait to trait (cf. Ch. 3) or from one condition to another. As is true in all these cases, if a solution is to be found it must be stated in terms of a specifically defined situation. Much of the controversy and confusion seems to have arisen from the attempt to go beyond the concretely established facts and discuss a sort of disembodied abstract "variability" which is expected to be independent of the particular situation in which it has been measured.

In any analysis of the effect of practice upon individual differences, it is necessary to ascertain at the outset what is meant by *equal practice*. If all individuals are permitted to practice for an equal period of time, the slower worker will be at a disadvantage since he will have received practice on less material than the faster individual. The use of an equal amount of material, on the other hand, places a handicap on the faster worker, who will necessarily have spent less time in learning the material than the slower person. The *amount limit* method, giving the advantage to the initially poorer individual, favors a decrease in variability with practice, whereas the *time limit* method favors an increase.

Each method answers a somewhat different question. The best criterion for choosing between the two seems to be a practical one. Equal training, as the term is used in everyday life, usually refers to equal time spent in training. When a person takes a "course" in music, or golf, or Spanish conversation, he is given a specified *number* of lessons, each of the same duration. No adjustment is made for the fact that during that period the number of times a piano key is touched or a golf ball is hit, or the number of words spoken differs widely from one individual to another. The time limit method would thus appear to be preferable, but either may of course be used. The important point is to take cognizance of which method was used, when interpreting the results.

A second problem which confronts us is the choice of a *measure of progress* to be employed. In Table 6 are illustrated three alternative ways of reporting the *same scores* obtained by two subjects, A and B, with the time limit method. In this table, A represents an initially faster worker and B an initially slower one. It will be noted that the

relation between the gains of the initially better and poorer subjects differs with each type of measure employed. When the scores are expressed as amount of work done per unit of time, the gain of the better subject appears larger than that of the poorer one. This will tend to make variability increase with practice. If, on the other hand, these same scores are expressed as time per unit of work, the slower individual will seem to gain more.

TABLE 6 *Various Ways of Expressing the Effects of Practice*

1 *Amount Scores*	*Subject*	*Number of Items Completed during a 1 Minute Trial*		*Gain in Items per Trial*
		First trial	Last trial	
	A	20	30	10
	B	12	20	8
2 *Time Scores*	*Subject*	*Average Time in Seconds to Complete One Item*		*Gain in Time per Item*
		First trial	Last trial	
	A	3″	2″	1″
	B	5″	3″	2″
3 *Time Saved per Trial*	*Subject*	*Gain in Items per Trial*	*Time Initially Required for Each Item*	*Gain in Time per Trial*
	A	10	3″	30″
	B	8	5″	40″

This apparent contradiction becomes intelligible if we realize just what time and amount scores are measuring. Since the slow worker requires more time on each item, *for every additional item which he completes in the later stages of practice, he will be saving much more time than the faster worker*. Thus, if it took the slow worker, B, 5 seconds to complete one item at the beginning of practice and if he can complete 8 more items after practice than he could before, he has gained the equivalent of 8 × 5 or 40 seconds per trial. The faster worker, A, on the other hand, added 10 items to his score, but he only required 3 seconds per item at the outset, so he has gained 10 × 3 or 30 seconds (cf. method 3, Table 6). The gain in time *per item* (method 2, Table 6) favors the slower worker even further, since it

does not take into account the fact that during any one trial this unit gain in speed is manifested *more often* by the faster than by the slower worker, the former completing more items.

It is apparent, then, that the problem of practice and variability must be further defined in terms of the measure of progress employed. If a choice is to be made among the various measures, amount scores will prove more serviceable because of their wider applicability. In a "speed" test, amount scores can be employed interchangeably with time scores. In a "power" test, however, in which the items are arranged in an order of progressively increasing difficulty, a time score would be meaningless. If, for example, in a 30-minute test consisting of 10 problems, all the subjects attempt all the problems but the number of correct solutions ranges from 1 to 10, it would be absurd to report that the average time per problem ranged from 30 minutes to 3 minutes. The better subjects did not necessarily work any faster than the poorer subjects, since all members of the group tackled all the problems.

A third problem pertains to the *inequality of units* in different parts of the scale. In many of the tests in which the items are arranged in increasing order of difficulty, the successive items do not progress by equal increments of difficulty. Frequently there are larger "gaps" between adjacent items at the extremes of the scale than there are between items near the center. Or there may be a relative scarcity of items at the upper end only, or at the lower end only. Such an unequal distribution of items would affect the meaning of differences in total scores at different parts of the range.

Let us assume, for example, that in a particular test the successive items are closer together in difficulty at the low end of the scale and farther apart at the upper end. An individual at the low end of the scale who obtained, let us say, an initial score of 16 items correct might very easily raise his score to 24 in the course of practice, thus apparently gaining 8 points. Another individual, near the upper end of the scale, who began with a score of 35, might achieve a final score of 40, thus gaining only 5 points. In this illustration, the initially poorer person makes a larger gain in raw score than the initially superior performer. If this occurred consistently, individual differences in raw score would decrease with practice, the members of the group being more closely alike after practice than before. In terms of equal-unit scores, however, the individuals at the low end of the distribution may have been improving much less than those at the upper end, since

each raw-score unit at the low end of the scale corresponded to a smaller ability difference than did a raw-score unit at the upper end. An 8-point gain from an initial score of 24 might actually represent less improvement than a 5-point gain from a score of 35. Inequality of units might thus lead to a completely erroneous conclusion regarding the effect of practice upon individual differences.

The changes in work method which often occur in different stages of practice are very likely to affect the relative distance between successive score units. If, for example, progress beyond a certain score requires a more complex organization of simple activities, or the development of a more efficient procedure, then score units at this point probably represent larger steps in a scale of difficulty level. Shifts in size of raw-score units may also occur in tasks in which a "physiological limit" is rapidly approached. This is often true in motor tasks and in many tasks in which speed is of primary importance. In such cases, physiological or structurally imposed limitations may make progress beyond a certain point impossible. As this point is approached, it becomes increasingly difficult to improve one's score; the successive score units thus correspond to progressively larger differences in difficulty level. The same effect occurs when progress is artificially limited by the test ceiling. If this ceiling is too low for the subjects being tested, it will have the effect of artificially reducing individual differences in the course of practice, since everyone's progress is arbitrarily cut short at a relatively low level, although a number of individuals could have advanced much farther.

Finally, a fourth consideration concerns the use of *relative or absolute measures of variability* in analyzing practice data. When absolute measures are used, such as the standard deviation, or gross gains made by initially high and low individuals or groups, variability tends to *increase* with practice. When, on the other hand, relative measures are employed, such as the coefficient of relative variability,[7] or some measure based upon relative or percentage gains, then variability *decreases* with practice in most cases. The fundamental objection against the use of relative measures has already been discussed in a previous chapter (cf. Ch. 3). It was there demonstrated that, since scores on most current psychological tests are not measured from an absolute zero point of performance, any ratios or quotients computed with such

[7] $\frac{100\,SD}{\text{Average}}$.

scores may be entirely misleading; the addition of a few easy items at the lower end of the scale might completely reverse the relationship between the obtained values.

Thus it would seem that absolute measures of variability are preferable for a purely negative reason, if for no other. Since relative measures are ruled out by the use of arbitrary starting points in the tests, no alternative is left. We may, however, inquire more directly into the logic of using absolute or relative measures in practice experiments. The argument in support of relative measures is that, since the numerical size of scores changes in the course of practice, the scores are not expressed in the same units throughout and hence absolute measures will not be comparable from trial to trial. Through chance alone, the argument runs, absolute variability will *increase* when the size of scores increases and *decrease* when the scores decrease, such changes being therefore of the nature of a statistical artifact.

It is perfectly true that, other things being equal, numerically larger scores will exhibit greater variability. Obviously, if the standard deviation of a distribution of time scores is 10 minutes, the standard deviation of the same scores expressed in seconds will be 600. For the same reason, the standard deviation of the number of A's cancelled in one minute cannot be compared directly with that of the number of additions performed during an equal period, since the latter scores would be much smaller.[8]

This type of argument does not necessarily hold, however, when the same test is given to different groups or to the same group under different conditions, such as before and after practice. Let us suppose that the average score of a certain group I on an intelligence test is 25 points and that of group II, 50 points. It does not necessarily follow from this difference in averages that group II will have a larger standard deviation than group I. In fact, the opposite might very likely be the case. If group I, for example, consisted of unselected third grade public school children and group II of superior sixth grade children in a private school, the latter would probably have a lower standard deviation. Similarly, one could assemble without too much difficulty two groups of men whose average heights were 64 and 72 inches respectively, but whose standard deviations were 8 in both cases. It would be quite absurd to insist that the taller group is "actually" less variable in height than the shorter, or to suggest that the inches used

[8] Cf. Ch. 3 for a fuller discussion

in measuring height had in some mysterious fashion changed in value from group I to group II. It would seem that the comparison of variability before and after practice is more similar in principle to the above examples than to the measurement of variability in different tests. The use of absolute measures in this connection is therefore justifiable.

TYPICAL EXPERIMENTAL FINDINGS ON PRACTICE AND VARIABILITY

The theoretical analysis given in the preceding section would lead us to expect opposite results when different techniques are employed to measure the effects of practice upon individual differences. The results actually obtained completely confirm the various theoretical expectations which have been outlined. Thus time or error scores show a decrease in absolute variability with practice (cf. 25); amount scores show an increase (cf. 2, 19, 20). Relative variability, i.e., the extent of individual differences expressed in relation to the level of performance at different stages of practice, nearly always decreases with practice (cf. 17, 19, 20). This simply means that the increase in individual differences with practice is not proportional to the increase in level of performance.

When the question of practice and variability, as studied by different investigators, is reformulated in comparable terms, the discrepancies among the results of different investigations disappear (cf. 2, 17). A meaningful answer to this question *can* be given, if the question is stated in specific terms.

An investigation by Anastasi (2) illustrates the procedure and findings of studies on practice and variability. For the reasons given in the preceding section, it was decided to define equal practice as *equal time spent in practice* and to express scores in terms of *amount done per unit time*. The scores on each trial of each of the tests were transmuted into an *equal-unit scale* of difficulty. The extent of individual differences at different stages of practice was measured by the standard deviation, a measure of *absolute variability*.

Four groups, each comprising from 114 to 200 college students, were given continuous practice in one of four tests. The tests included A-Cancellation, Hidden Words,[9] Symbol-Digit Code Learning, and

[9] Subjects were to underline all four-letter English words which were "hidden" in a page of pied type.

Vocabulary Learning[10] The practice consisted of 15 4-minute trials in Hidden Words and 20 2-minute trials in each of the other tests, a different group of subjects being employed for each test. The average scores and standard deviations of the scores on each trial are reproduced in Table 7.

TABLE 7 *Averages and Standard Deviations of Scores on Successive Trials* *

(From Smith [illegible])

Trial	*Cancellation* Average	SD	*Symbol Digit* Average	SD	*Location* Average	SD	*Hidden Words* Average	SD
1	40.63	6.78	41.15	7.58	39.06	6.84	43.58	6.94
2	44.99	6.42	47.63	7.38	46.70	6.03	44.63	6.90
3	47.00	6.60	52.69	7.30	45.22	6.95	49.00	7.52
4	48.00	6.52	54.57	8.04	47.74	5.88	51.25	7.74
5	50.75	6.60	57.90	7.94	49.19	6.56	54.49	7.86
6	50.30	6.68	58.63	8.34	48.80	6.78	55.12	8.24
7	51.68	6.62	61.02	8.66	52.06	7.22	58.18	9.28
8	52.74	7.04	62.27	8.44	48.97	7.89	60.40	8.90
9	53.06	7.28	63.79	8.08	51.59	7.16	61.30	9.10
10	55.83	7.24	64.52	8.36	52.50	8.34	64.10	10.22
11	54.70	7.24	65.22	7.94	53.08	8.90	62.19	10.46
12	55.08	7.22	65.70	9.40	55.35	8.10	63.26	10.96
13	56.09	7.70	67.04	8.06	54.54	7.98	67.02	11.36
14	55.50	7.12	67.51	8.40	54.74	7.26	68.47	12.96
15	57.88	7.54	67.78	8.72	56.02	8.49	69.28	11.44
16	56.67	7.70	69.12	9.78	56.48	8.46		
17	57.01	7.32	68.16	8.92	57.83	8.59		
18	57.62	7.58	68.81	8.80	56.63	9.13		
19	57.08	7.36	69.17	8.40	56.97	8.89		
20	59.60	7.88	70.07	9.98	59.28	8.87		

* The scores on all these tests were transmuted into an equal unit scale and are thus directly comparable from one test to the other.

It will be readily seen that the standard deviations rise with practice in every test. It was also found that individuals tend to maintain the same relative standing in the group in the course of practice, the correlations between initial and final scores of the same subjects being con-

[10] Subjects learned by the method of paired associates a vocabulary of nonsense syllables; the test is similar to code learning but more difficult.

sistently positive and usually high For the four tests, these correlations were:

Cancellation	6725	Vocabulary	5073
Symbol Digit	2981	Hidden Words	8239

Such correlations indicate a tendency for the individual who is best in the group at the outset to remain at the top after practice, for the one who is lowest to remain at the bottom, and so on This is commonly found to be the case in all experiments on practice

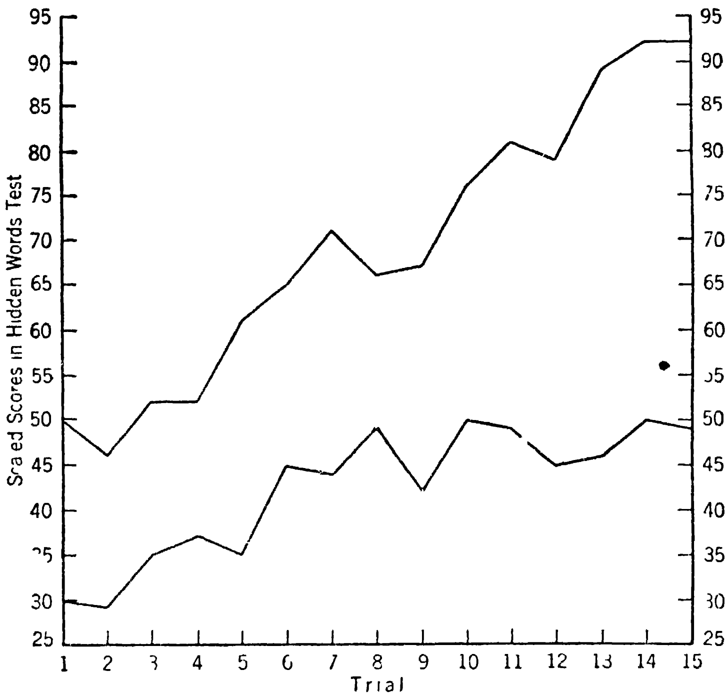


Fig. 45. Learning Curves of Two Subjects Illustrating Divergence with Practice (Unpubl data from investigation of Anastasi 2)

Both the tendency to maintain the same relative position during practice and the increase in absolute variability are illustrated graphically in Figure 45 This shows the *learning curves* of two subjects on the Hidden Words test The subjects were selected near the extremes of the distribution, the difference between their initial scores being very large It will be noted that the curves do not at any time cross and that

they diverge with practice, the difference between the two individuals being much larger on the fifteenth trial than it was on the first.[11]

THE STUDY OF PRACTICE AS AN APPROACH TO THE HEREDITY-ENVIRONMENT PROBLEM

Some writers have seen in the data on practice and variability certain implications for the underlying question of the relative influence of hereditary and environmental factors. It has been argued that if individual differences in performance *increase* with practice, they can be attributed to hereditary differences, while if they *decrease*, they must have resulted, at least in large part, from inequalities of past training and environmental stimulation. Probably the first explicit formulation of this hypothesis was made by E. L. Thorndike in 1908. In an article appearing in that year, he wrote:

> Experiments in practice offer evidence concerning the relative importance of original nature and training in determining achievement. In so far as the differences among the individuals in the ability at the start of the experiment are due to differences of training, they should be reduced by further training given in equal measure to all individuals. If, on the contrary, in spite of equal training, the differences amongst individuals remain as large as ever, they are to be attributed to differences in original capacity (22, pp. 383-384).

More recently, Thorndike (24) conducted an analysis which represents a specific application of this hypothesis. The data consisted of the scores obtained by several groups of high school and college students on Cooperative Test Service examinations in English, Latin, and several modern languages. The number of students in each group varied from 50 to 2767. For the purposes of such an analysis, Thorndike considered the time spent in learning each of these three subjects as the contribution of environment to achievement on the corresponding examinations. Similarly, he regarded individual differences in score within the groups which had had the same length of training as being the result of heredity. He then found the AD[12] of the combined

[11] Different relations may be obtained if individuals are at different stages of the practice curve at the starting point, i.e., if they have had differing amounts of relevant practice prior to trial 1.

[12] Average deviation, being the average of the differences between each individual's score and the group mean. This is a simpler and cruder measure of variability than the SD.

groups, throwing together those who had received different lengths of training. This AD would obviously reflect the combined influence of heredity and environment. From this value he deducted the AD of the specific training groups, taken separately. Since the total AD was reduced by only about 25%, on the whole, by this subtraction, Thorndike concluded that this per cent represented the net contribution of environment to individual differences in the scores on these examinations. The major influence he thus attributed to hereditary factors.

A critical evaluation of this study is given by Hamilton (13), in terms of both statistical methodology and interpretation. First it is demonstrated mathematically that the AD does not lend itself to the analysis by subtraction which Thorndike employed. The residual AD does not correspond mathematically to the contribution of practice, or length of training, as Thorndike had assumed. Secondly, Hamilton calls attention to the fallacy of attributing to heredity all individual differences which remain when length of training is constant. Such an assumption would ignore all the differences in the students' previous experience which might result in differences in motivation, study habits, previously acquired information and skills, and the like. The students' performance in the courses, and consequently in the examinations, would obviously be affected by such antecedent environmental factors. To this should be added the well known fact that registration in the same course does not signify the same amount of time spent in learning the subject on the part of different students!

In a study designed to avoid the pitfalls discussed above, Hamilton (13) gave groups of from 22 to 28 fifth grade school children 20 trials of each of three learning tasks, viz., artificial language, symbol-digit substitution, and "making gates." By more refined statistical analyses,[11] Hamilton demonstrated that the amount of practice, i.e., the number of trials which the subjects had had at any one stage during the experiment, was a more potent determiner of achievement than Thorndike's results had suggested. For example, when performance on trials 1 and 2 was compared with performance on trials 19 and 20, the "practice effect" far exceeded the combined effect of all other factors causing individual differences. Table 8 shows the relative contributions of present practice on the one hand, and of other, residual factors on the other, in this particular comparison.

[11] Analysis of variance and intraclass correlation. The variance of a distribution is the average of the squared differences of each score from the group mean, i.e., SD^2.

TABLE 8 *Analysis of the Contribution of Practice to Individual Differences*

(From Hamilton 13, pp. 32-34)

Test	*Per Cent of Individual Differences Attributable to Amount of Experimentally Administered Practice*	*Per Cent of Individual Differences Attributable to Other, Residual Influences*
Making Gates	84.15	15.85
Symbol-Digit	55.65	44.35
Artificial Language	71.43	28.57

That the conclusions reached by Thorndike in the earlier study were in part the result of faulty statistical methodology was demonstrated when Hamilton recomputed some of Thorndike's figures by the present method. Hamilton found that the per cent of the variability attributable to length of training was much greater than had been obtained in Thorndike's analysis. She concluded not only that practice plays a much greater role in determining individual differences in achievement than had been suggested by Thorndike, but also that no single estimate of its relative contribution can be given. She points out that the proportional contribution of practice depends upon: (1) the stage of the learning curve at which individual differences are measured; (2) the amount of practice which intervenes between the trials being compared; (3) the heterogeneity of the groups in regard to other relevant characteristics; and (4) the kind of task or skill under consideration.

The second and third of the conditions cited by Hamilton will be recognized as a specific application of some of the points implicit in the concept of *interaction*, as discussed in Chapter 4 in connection with the general problem of heredity and environment. It was there pointed out that estimates of proportional contribution are inconsistent with our knowledge of the operation of hereditary and environmental factors. A different estimate will be obtained in groups which vary in their environmental or hereditary heterogeneity. In a group having highly similar environment, hereditary factors would have a larger weight in determining individual differences. Conversely, in a group highly similar in heredity, environmental factors would exert a relatively greater effect in the development of individual differences. It

should be noted that the same relationship holds when determining the relative weights of two different environmental factors or two different hereditary factors. Thus in the above experiment, if we wish to compare the relative contribution of immediate practice, i.e., number of trials or length of training, with the contribution of antecedent factors (environmental and hereditary), the estimate would vary as either present practice or prior conditions vary. For example, in comparing individuals all of whom have had exactly the same number of practice trials, the contribution of immediate practice would obviously be zero. Similarly, if the scores on trials 15 and 16 are compared, the role of practice in producing score differences will appear to be relatively small. When, on the other hand, we compare groups which differ markedly in the number of trials which they have had, then the role of practice in individual differences will be large. If we are interested in discovering how far practice may account for individual differences, then we should obviously give practice the opportunity to operate, by comparing individuals who differ conspicuously in amount of practice.

In concluding the present section, one other implication of practice and variability studies ought to be examined. It is sometimes argued that when subjects undergo a prolonged period of equal training, the differences in their past experience with the given task are thereby wiped out. This assertion is open to question. The influence of environmental factors upon the development of the individual is ordinarily *cumulative*. If one individual's past experience has made him more proficient than another in a certain task, we should expect him to be better fitted to profit from instruction for that very reason. Susceptibility to training can itself be environmentally determined, and if so determined there is no reason to assume that it will disappear with additional training.

The individual who has been handicapped by a "poor" environment may lack the necessary intellectual tools to profit from instruction. Thus, had the Wild Boy of Aveyron (cf. Ch. 6) and a boy of the same age from a middle-class English home been put through an identical one-year course in the reading of French, the differences in their abilities to read that language would have been far greater at the end of the course than at the beginning. Similarly, investigations on the acquisition of motor skills (cf. 14) have demonstrated that individuals who have been taught more efficient work methods not only have a head start, but with continued practice gain progressively more than others

using less efficient methods. It is obviously unnecessary to assume a hereditary basis for individual differences in order to account for the increase in variability in such examples The more the individual has learned in the past, the more he will be able to learn in the present To use a rather crude analogy we might say that practice does not add to the individual's ability, but multiplies it

CHAPTER

8

Schooling and Intelligence

A LIVELY CONTROVERSY has centered about the effects of schooling upon intelligence test performance. The divergent conclusions reached by different investigators have resulted at least in part from inadequate clarification of underlying concepts—a fact which has led to prolonged critiques, replies, rejoinders, and counter rejoinders in the journals, with the participants being no closer at the end than they were at the outset. Scores of investigations have been conducted, some emphasizing and others minimizing the role of schooling in the development of intelligence. When the mass of available data is sifted, no very startling discovery regarding heredity and environment emerges. As an experimental approach to the problem of heredity and environment, the study of the effects of schooling leaves much to be desired. Some of its limitations will be considered in the analysis of the theoretical implications of this method, to be given in the concluding sections of the present chapter.

If these studies have contributed little to the sum total of our knowledge regarding the factors operative in behavior development, they have nevertheless indirectly stimulated a thoroughgoing evaluation of practices commonly followed in mental test studies. In the course of the controversy, attention has been focused upon the methodological requirements of such investigations. Needed cautions in the interpretation of statistical data have been clearly expounded, and rigid standards for the control of conditions have been set forth. In looking over the critical literature concerning the studies on "schooling and the IQ," one cannot escape the impression that higher standards were demanded than had heretofore been commonly applied in most mental test studies, on any topic In their zeal to counteract a too sweeping

generalization or a premature publicizing of results, the critics sometimes outdid themselves. The net outcome, however, has been a positive contribution to the development of sound methodology in mental test studies and an increasing awareness by investigators in this area of the need for experimental controls and careful evaluation of statistical findings.

At the core of the controversy are a group of studies dealing with the effects of nursery school attendance upon the IQ. It is with this *preschool level* that the largest number of investigations and much of the discussion have dealt. Nevertheless, a number of studies which followed essentially the same procedure have been conducted at higher educational levels, *from the elementary school through college*, and these will also be included in the present chapter. A few studies concerned with the influence of *special educational techniques*, such as specialized courses of training or specially designed curricula, will also be considered. The latter studies are somewhat related in general approach to the *coaching studies* covered in the preceding chapter. They differ, however, in that the training is much farther removed from the actual test content and was not designed with reference to the test.

Investigations on the effects of schooling also have certain features in common with studies on the effects of various *institutional environments*. The latter will be discussed in Chapter 11, in conjunction with the investigation of foster home environment. All the studies treated in the present chapter deal specifically with schooling or training, as distinguished from the more general factors operative in the "home" or "living" environment of the individual.

THE EFFECT OF SPECIAL EDUCATIONAL TECHNIQUES

A few investigators have been interested in the possibility of raising the intellectual performance level of dull or feebleminded subjects by means of specially designed, intensive programs of training. In a study by Kephart (25, 26), 16 boys living in a single cottage in a training school were given special instruction for a period ranging from six months to nearly three years in individual cases. The aim of the program, according to the author, was to stimulate constructive activity and to encourage ingenuity, initiative, and original planning. Concrete materials, social situations, and abstract problems were included in the

training. Among the latter were problems involving the recognition of absurdities in stories, a task which has much in common with some of the Stanford-Binet tests. At the beginning of the experiment, the age of the group ranged from 15 to 18, with an average of 16–6 (i.e., 16 years and 6 months). Initial Stanford-Binet IQ averaged 66.3 and ranged from 48 to 80. At the end of the experimental period, the mean IQ had risen to 76.4. Individual gains ranged from 2 to 22 IQ points. All subjects gained upon retesting except one, who lost 3 points. Half of the group gained 10 or more points; one fourth gained 15 or more. Control data were obtained within the same group, as well as in an equated group. Tests given to the experimental group over an eight-month period prior to the initiation of the special training program showed a mean rise of only 2.3 IQ points, the individual changes ranging from −7 to +18. A control group of 26 boys of the same age and in the same institution as the experimental group was tested over the interval in which the experimental group participated in the training program. The mean gain of this control group on the retest was 1.9, with a range from −10 to +15.

An extended project on the education of mentally retarded children reported by Schmidt (45) has aroused a storm of controversy. In this study, 254 boys and girls between the ages of 12 and 14, who had been referred to special classes, were put through a three-year educational program especially designed for them. The average initial Stanford-Binet IQ reported for this group is 52.1, with a range from 27 to 69. The subjects were tested periodically with intelligence, educational achievement, and personality tests during the three-year training period, as well as during a five-year follow-up after the completion of the experimental program. The degree of progress in all aspects of behavior reported in this study far outstrips that found in any other investigation to date.[1] At the completion of the investigation, a mean

[1] In view of its scope, duration, and wealth of observations, this study deserves serious consideration. At the same time, the investigator herself calls attention to the wide divergence of these results from conventional professional opinion and points out the need for independent verification. A critical analysis of this study, together with suggested reasons for its discrepant results, will be included in the general evaluation at the conclusion of the present section.

An adverse critique by S. A. Kirk, together with a reply by Schmidt, appeared in *Psychol. Bull.*, 1948, 45, 321–343. Since the reviewer did not have access to all of Schmidt's cases, and since his sources of data show certain internal inconsistencies, it is difficult to draw any conclusion from this exchange of comments. At this stage the safest conclusion is that the study offers valuable leads for further research. The reader should also consult the critical review by F. L. Goodenough in *J. Abn. Soc. Psychol.*, 1949, 44, 135–140.

gain of 40.7 IQ points was observed; 80.7% of the subjects made gains of 30 or more IQ points and 59.6% gained 40 points or more. The larger part of these gains occurred during the three-year experimental period, although in the course of the subsequent five-year follow-up the IQ's showed *continued gains,* rather than dropping toward the initial level.

The progress in educational achievement reported by Schmidt is equally remarkable. Although the average educational performance at the beginning of the experiment fell within the first grade, by the completion of the three-year program it had reached approximately fifth grade level. Moreover, 79 subjects transferred to the regular elementary school either to qualify for immediate graduation from the eighth grade or to complete the elementary school course in regular classes. During the five-year follow-up period, a large number continued their education in technical, business, or avocational courses, and 27 of the original group had graduated from high school by the termination of the study. Data on subsequent occupational history, socio-economic status, community activities, and the like during the follow-up showed the group to have made a very satisfactory adjustment.

As a control group, Schmidt employed 68 children, also enrolled in special classes for the intellectually deficient but not participating in the experimental program. The control group was approximately equated with an experimental sub-group of 64 cases in initial IQ, educational achievement, and chronological age. The mean gain of this experimental sub-group was 23.8 IQ points, while the control group lost an average of 3.6 points during the same period. Marked differences in educational progress and in subsequent vocational and social adjustment were likewise found between these two groups.

Both of the above studies suggest that special training may exert considerable influence upon intellectual development. The reverse conclusion was reached by two other studies conducted on dull-normal and normal subjects. In one of these (42), 111 dull-normal children were given the Stanford-Binet before and after a two-year period in a school offering an "experience curriculum." The authors report that the curriculum was especially planned to stimulate intellectual activity among slow learners at this ability level. Pupil interest is described as very high and truancy was virtually eliminated during this program. It was originally planned to admit to this course only children with

IQ's between 75 and 90, but the sampling actually included 10 cases between 60 and 74, and 6 between 96 and 104; the latter were admitted because they represented decided school failures, despite their near-normal IQ's. The initial mean IQ of the sampling tested was 85.12. Initial age ranged from 5–8 to 12–3. The mean change in IQ during the experimental period was slight and similar, in both direction and amount, to retest changes found by Terman and others when the Stanford-Binet is readministered with no special interpolated experience.

Similarly, the Stanford-Binet IQ's of 141 children who completed the fourth grade of a demonstration school were not increased by participation in what is described as a "rich and vital school curriculum" (28). The mean initial IQ of this group was 109, with a range from 89.0 to 134.6. Fifty-one of these children, who were retested over a four-year period, showed a mean IQ loss of 1.53 points. A group of 74 retested after three years gained an average of 1.48 points, and 89 children retested within two years made an average gain of .06. None of these differences is statistically significant.

A somewhat different approach is illustrated by an experiment (54) in which 30 college sophomores were given six weeks of training in general semantic methods. Scores on the Detroit Intelligence Test, Advanced Form, rose an average of 36 points during this period, in the experimental group. The control group, which had received no such training, showed an average rise of 6 points. In terms of the national norms for this test, the gain made by the experimental group represents a rise from the 62nd to the 96th percentile.

Mention may also be made of the results achieved with the army's Special Training Units during World War II (6). Through an intensive 12-week course of instruction in these units, men who had been illiterate or of very limited education were brought approximately to the fourth grade elementary school standard in reading, language expression, and arithmetic. At the same time, their performance on the AGCT rose from Grade V, the lowest army grade, to Grade IV or even higher. Approximately 85% of the men selected for such training succeeded in reaching such standards. Had the initial classification of these men been regarded as an index of their "native intellectual capacity," without reference to their educational and other experiential limitations, the possibility of "raising" them to Grade IV level would have been overlooked.

In evaluating any of these studies, a fundamental question is: How broad or how narrow was the effect of the particular training which was furnished? It is not at all surprising that the results should differ with the nature of the training, the degree of similarity between the trained functions and the functions sampled by the tests, and possibly the initial intellectual level of the subjects.

In connection with the positive findings reported by some of these studies, one may ask to what extent the improvement was limited to functions closely similar to the tests, and to what extent other intellectual behavior had also improved. In so far as the educational achievement of the subjects in the Schmidt study also showed marked gains, and in the light of the subsequent vocational and social adjustment of this group, the area of improvement appears to be considerably broader than that of the test. This is also true of the results obtained in the Army Special Training Units. It would be misleading, however, to assume that the rate of development of all behavior functions had been accelerated by such training, or to speak of improvement in some mysterious "underlying capacity." What is affected is observable behavior, and the breadth of behavior so influenced can be empirically determined for each type of training.

Applying the same analysis to the two studies which yielded negative results, we find, first, that such training procedures as are subsumed under the "experience curricula" and the "rich, vital curricula" of progressive education seem not to affect appreciably the IQ of most normal or borderline children. Taken as it stands, this finding is not too unexpected. Such curricula generally emphasize interest, individual initiative, practical applications, and a number of similar features which may help the general adjustment and achievement of the individual, both in school and out. But such instruction is not oriented toward improving the type of behavior functions which are predominantly sampled by most intelligence tests. Among the latter functions we may note, for example, abstract verbal and numerical ability, memory, attention to details, speed of routine work, and following directions minutely and without hesitation. Whether other intelligence tests should be devised to sample different behavior functions, or whether the progressive curricula cover the most desirable functions to be developed in any one group, is of course entirely beside the point. What is relevant in the present connection is the fact that different

curricula or courses may differ widely in their coverage of the type of behavior functions sampled by intelligence tests. When the problem is viewed in this light, the results of the studies which have been cited, although widely divergent, need not be regarded as inconsistent or contradictory.

In the Schmidt study, which reports the most conspicuous effects of training, the experimental program was very broad in its coverage. Although it, too, was concerned with the stimulation of pupil interest, a multiplicity of procedures which might lead to intellectual improvement were included. For example, the attention given to the development of effective work and study habits and to the attainment of minimum levels of performance in reading and language usage may account in part for the continuance of improvement after the termination of the experimental period. These skills provided the necessary tools for further progress. The care taken to adapt instruction to the specific needs and deficiencies of each individual may also have contributed to the effectiveness of this training.

On the other hand, it would undoubtedly be rash to generalize these results to all cases of intellectual backwardness. Selective factors probably operated to make the particular sampling of this study more susceptible to rapid improvement than would be the case among intellectually retarded subjects in general. The group as a whole came from very inferior socio-economic backgrounds, where opportunities for normal behavior development were poor—a condition with which the ordinary elementary school could not adequately cope. There is the further likelihood that the educational and consequent intellectual development of a number of these subjects was initially hampered by sensory defects, poor health, and language handicap. All these conditions could be—and probably were—largely remedied in the course of the experimental period, which would account for some rapid gains in educational performance as well as in IQ. The subjects as a group were likewise socially immature and poorly adjusted emotionally at the outset. Marked improvement in these respects is reported in the course of the experiment. Such improvement would in turn affect intelligence test performance both directly and indirectly: directly through greater alertness, interest, and cooperation during the test itself, and indirectly through an increase in the effectiveness of learning in general and in the acquisition of those skills which are sampled

by intelligence tests. It is interesting to note in this connection that those individuals showing the greatest improvement in emotional and social adjustment also tended to show the greatest gains in IQ. Thus the correlation between improvement in Stanford-Binet IQ and in Bernreuter B1N score (emotional adjustment) was .923, and that between IQ gain and gain in the Vineland Social Maturity Quotient was .874.

In the light of these considerations, the results reported by Schmidt are probably not so startling as they might appear at first sight. A conclusive evaluation of this study, however, would require more information than is provided by its author. Especially would it be helpful to know more about the detailed procedures employed both in the training program and in the testing. The presence of a number of minor arithmetic errors and inconsistencies in the published results also suggests an unfortunate carelessness in reporting data.

In summary, it is apparent that both the nature of the training and the nature of the subjects determine the degree to which intellectual performance level can be raised by training. Much more research is needed to ascertain the relative effectiveness of different types of training for different individuals, as well as the relationship between subject characteristics and susceptibility to training in general.

STUDIES ON PRESCHOOL ATTENDANCE

Over fifty investigations have been conducted to determine what effect, if any, preschool attendance at a kindergarten or nursery school has upon the child's IQ. A few studies (e.g., 8, 34, 35, 48) give only the intelligence test scores of a nursery school group before and after a period of preschool attendance. In such studies it is impossible to determine how much of the change in score may result from retesting or from the time of the year when the tests are given. A control group is essential for this purpose. Another group of studies (e.g., 29, 39, 40, 41, 56) report only the relative performance in the first grade, or at subsequent scholastic levels, of two groups, one of which had attended preschool while the other had not. The difficulty with this procedure is that nursery school attendance may be—and probably is—itself selective. Even when the groups are equated in parental education and occupation, as well as in other broad categories, selection may

have occurred within these categories. For example, among families with the same educational, occupational, and socio-economic level, those parents who enroll their children in nursery school may still have differed in intelligence, personality characteristics, interest in the children, and other subtle and inconspicuous ways. A local type of selection may also occur in certain nursery schools which offer special remedial services. In such cases, the children with defects of speech, personality, etc., are probably more likely to be sent to the preschool. Corroborative evidence for the operation of such selective factors is furnished by some of the investigations (cf. 29, 56).

The most direct analysis of the effect of nursery school attendance is based upon the intelligence test scores of a nursery school group before and after a period of preschool attendance, together with the scores of a matched control group tested and retested over the same interval. This is the procedure which has been followed by the majority of investigators. In some of the studies, successive retesting at intervals within the preschool period permits the investigation of cumulative effects and the determination of the course of the changes throughout the period. Moreover, a follow-up of the experimental (or preschool) group and the control group for several years permits a study of the permanence of the effects observed.

In their interpretations, most investigators have aligned themselves definitely on one side or the other of the controversy. Some lay great emphasis upon the differences which have been found in favor of the nursery groups. Others stress the smallness of such differences and their complete absence in some of the groups. The findings vary, to be sure, for a variety of reasons to be considered in a later section. But the interpretations vary more sharply than do the results. The data appear to fall, not into two categories, the pro and the con, but rather into a continuum of slightly varying effects, which may be related to the conditions of the investigations.

In a summary of about fifty studies on nursery school children by different investigators, Wellman (61) reports the results obtained with several intelligence tests. The largest number employed some revision of the Binet scales (Kuhlmann-Binet, 1916 or 1937 Stanford-Binet). When the results of these studies were combined, the mean gain by 1537 children in 22 nursery groups was 5.4 IQ points; the mean gain by 597 control, non-nursery children in 14 groups was 0.5. Mean gains

of over 6 points were reported for 50% of the nursery groups and 14% of the non-nursery groups [2]

The Merrill-Palmer Scale shows a larger influence of nursery school attendance, probably because of the greater similarity of its content to nursery school activities The mean gain of 267 children in 7 preschool groups on this test was 14 5 IQ points, that of 73 non-preschool children in 4 groups was 6 7 A mean gain of over 10 points is reported for 5 of the preschool but only one of the non-preschool groups A few studies employed a number of other intelligence scales, such as the Minnesota, Gesell, and California Preschool Schedules On these tests both preschool and control groups tend to gain in mean score upon retesting, with no significant or consistent advantage of the preschool groups

The comparison of gains made over initial and subsequent intervals of nursery school attendance shows that in nearly every group the increases in score are cumulative Later gains are however, slight, and the evidence strongly suggests that the major improvement in intelligence test performance occurs during the *first few months* of nursery school attendance

Many of the nursery school studies covered in the Wellman summary were based upon a small number of cases, any study covering 10 or more children having been included in the survey In a number of the studies the conditions of the investigation or the analysis of data were such as to make evaluation difficult Among the more ambitious projects, from the viewpoint of number of cases, duration of the observations, and number of factors investigated are those conducted at the Universities of Iowa, Minnesota, and California.

Wellman and her collaborators at the University of Iowa have conducted an extended series of projects on the effects of nursery school attendance upon intelligence test performance (60) The principal data were derived from a total sampling of 652 children between the ages of 18 and 77 months, who were attending either the nursery school or kindergarten conducted by the university All were given either the 1916

[2] The statistical significance of the gains is not always reported in these studies On the basis of available data the studies seem to be about equally divided into those which meet the common criterion of a significant difference between final scores of nursery and non nursery groups and those which do not In the comparison of initial and final means of the nursery groups the differences are also significant in about one half of the studies and insignificant in the other half In the non nursery groups however significant improvement is rarely found

[3] Included in Wellman summary as are those reported in 3 5 12 24 43 65

Stanford-Binet or the Kuhlmann-Binet in the fall [4] and again in the spring [4] of each year of preschool attendance. The mean difference between fall and spring tests during the first year of attendance was a gain of 6.6 IQ points; the changes ranged from a gain of over 40 to a loss of over 30 IQ points. Slightly over half of the children showed a change of 8 or more points.

Within the total sampling, 228 subjects attended preschool for two years or more, 67 of these attending for three years or more. Analysis of these sub-groups indicated that the mean score rose during successive years of preschool attendance, but the gains became progressively smaller. Thus the two-year group showed a mean fall-to-spring rise of 7.0 and 3.8 points, respectively, during their first and second years. The three-year group gained an average of 7.7, 4.3, and 1.7, respectively, from fall to spring testing during each successive year.[5] That length of attendance at nursery school bears little relation to the amount of gain was also demonstrated by the absence of significant correlation between gain and number of days of actual attendance during the year, the latter ranging from 37 to 148 days for individual children. Moreover, no relationship was found between the exact length of time which had elapsed between any one individual's fall and spring testing and his gain in score.

No relationship was found between amount of change in IQ among the nursery school children and the occupational or educational level of their parents. It should be noted, however, that the group as a whole came from superior occupational and educational levels Had the spread of parental characteristics and of home environments been wider, some relationship might have been found between these characteristics and gain in IQ.

In order to obtain control data for the evaluation of the observed gains in score, Wellman compared 34 pairs of preschool and non-preschool children, matched in chronological age and initial IQ. Between the fall and spring testing, this preschool group gained an average of 7.0 points, while the control group lost an average of 3.9 points. The mean difference of almost 11 points between the two groups on the spring test was statistically significant.

Follow-up studies in elementary school, high school, and college have also been conducted on nursery and non-nursery groups in the Iowa project. In a frequently cited study by Kounin (27), the achievement of 22 preschool and 31 non-preschool children was compared during the

[4] The "fall" tests were given between August 1 and December 31, most of them occurring in October and November; the "spring" tests extended from March 1 to June 30, with the maximum concentration in April and May.

[5] The net gains over the total period are smaller than the sum of these annual gains, since a slight loss in mean score occurred during the summer months The "initial" score each fall was thus slightly lower than the "final" score of the preceding spring.

first four grades of elementary school. The two groups were approximately equal in initial IQ, the mean Binet IQ's for the period prior to preschool attendance being 118.3 and 117.7 for preschool and non-preschool groups, respectively. Achievement test scores in arithmetic showed no significant difference between the two groups; in reading achievement tests, slight and not very significant differences were found in favor of the preschool group. In school marks, no significant differences were noted during the first two years, but significant differences in favor of the preschool group appeared during the third and fourth school grades. The delay in the appearance of this difference might suggest that the preschool and non-preschool samples actually differed in intellectual level (through differences in home background or any similar factor other than nursery school attendance), but that such differences did not enter into the type of behavior sampled by intelligence tests at the preschool ages. Thus the initial equating of mean IQ's between the two samples would not be sufficient to rule out other relevant differences, and the later divergence in school achievement could not be conclusively attributed to the effect of nursery school attendance. The evidence presented by this study is rendered even more uncertain by the small number of cases involved in some of the comparisons, since by the end of the fourth school year the groups had shrunk to 10 preschool and 8 non-preschool children. A further difficulty is the fact that the groups compared at the upper grades were no longer equated in initial IQ, owing to the selective elimination of cases.[6]

In another study, Wellman (58) compared 29 preschool with 29 non-preschool subjects who had been matched on initial IQ and years of school attendance subsequent to preschool. The intelligence test scores of this group during the high school period yielded a negligible and insignificant difference in favor of the preschool group. Equally insignificant was the difference found during the college period between initially matched groups of 19 preschool and 19 non-preschool cases.[7] The high school and college groups overlapped, some of the same subjects being included in both groups; hence the results of the two comparisons cannot be regarded as independently corroborative. One can only conclude from these studies that no prolonged effects of nursery school attendance upon either intelli-

[6] To be sure, the differences in school marks in favor of the preschool group in the third and fourth grades remained when only individuals of approximately the same IQ's were compared, but this necessitated a still further reduction in the *number* of cases compared.

[7] The mean differences in *initial IQ* in favor of the preschool groups were 0.5 and 2.7 for the high school and college samples, respectively. The corresponding differences in percentile scores on the American Council Psychological Examination and the Iowa College Entrance Examination were 9.8 and 12.0. These two differences are 1.9 and 1.7 times as large as their respective standard errors, thus being quite insignificant. Moreover, the averaging of percentile scores in these data distorts the results somewhat and makes their interpretation more difficult (cf. Ch. 2).

gence test performance or school achievement have been satisfactorily demonstrated.

In another investigation conducted by the Iowa group (46, 62), two matched samplings of normal and borderline *orphanage children* were studied over a three-year period. The experimental group attended a preschool conducted at the orphanage; the control group did not. The two groups were matched in initial IQ (on either Kuhlmann-Binet or 1916 Stanford-Binet), CA, sex ratio, length of residence in the orphanage, nutritional status, and presence or absence of sensory defects. The principal finding of this study was that the mean IQ of the control orphanage children dropped during their institutional residence, while that of the preschool group either rose slightly or showed a negligible change. Preschool attendance in this group seems thus to have counteracted the relatively "unstimulating" environment of the orphanage. This difference between preschool and control children was observed at all IQ levels, although it was not significant throughout. The decrease in IQ in the non-preschool group was progressive with increasing length of institutional residence. Thus no substantial change in IQ was observed in this group over the shortest residence period studied (averaging 115 days). At the other extreme, children whose stay at the orphanage averaged 642 days lost an average of 16.2 IQ points. The effect of preschool attendance upon the IQ, on the other hand, was manifested early and showed little or no subsequent change with continued attendance.

Among the various criticisms which have been directed against this study, one or two are particularly relevant to the principal results cited above. It has been pointed out (15, 36), for example, that adequate matching of the preschool and non-preschool groups was not sustained throughout the experimental period, because of the removal of children for foster home placement and their replacement in the experimental groups by substitutes. As a result, the 47 children who at one time or another were in the preschool group averaged 3.4 IQ points higher than the 44 children in the control group. The preschool group had a mean initial IQ of 86.9, with a range from 65 to 163; the initial IQ's of the control group averaged 83.5 and ranged from 57 to 114.

In a re-analysis of the data (62), it was demonstrated that among the children with over 400 days of orphanage residence, those who had attended preschool differed significantly in final but not in initial IQ. Within the same residence group, the preschool children who had actually attended the preschool for less than half of the total number of days did not differ significantly in IQ from the control group at the end of the observation period. Those who had attended the preschool for more than half of the period, on the other hand, showed a clearly significant difference

from the control subjects. In the last-mentioned comparison, the preschool subjects gained on the average 6.8 IQ points, while the control lost 6.1 points. This analysis suggests that the preschool training did actually serve to raise the children's intelligence test performance. Such a conclusion, however, requires further corroboration because of the small number of cases involved in any one of the specific comparisons made. A closer matching of the control and preschool groups in initial IQ and age would also permit a more precise interpretation of the observed gains and losses. It should also be noted that certain other conclusions drawn from this study by the investigators are open to serious question and will be considered in connection with methodological problems in a subsequent section.

One of the most carefully controlled studies on the effects of nursery school attendance upon IQ is that conducted by Goodenough and Maurer (18) at the University of Minnesota. A total of 147 children who had attended nursery school from 40 to 575 days were compared with 260 children having no nursery school training. All subjects had been tested at least twice with either the Minnesota Preschool Scale or the Stanford-Binet, an interval of at least one year having elapsed between tests. The children as a whole were above average in parental occupation and in their own initial IQ's, which averaged close to 110.

A special precaution was to have the tests administered by examiners who had no knowledge of the child's previous test performance and, in the case of at least 80% of the children over 6, no knowledge of their previous nursery school attendance. Moreover, the examiners were not connected with the nursery school. Thus their degree of mutual acquaintance with the nursery group was no greater than with the non-nursery group. It is pointed out that such acquaintance might have had a two-way effect, through both the examiner's and the child's attitude, in raising the score of the preschool group. Special efforts were also made to secure conditions of maximum motivation. Children were not tested on their first visit if they exhibited negativistic behavior. If a continued uncooperative attitude still rendered testing unsatisfactory, the case was excluded from the study.

The mean gains obtained in the retesting with the Minnesota Preschool Scale after one, two, and three years of nursery school attendance are given in Table 9. The advantage in favor of the nursery school group, according to these data, is either negligible or lacking. It was further shown that amount of nursery school attendance bore no relation to subsequent rise in IQ. The correlation between Stanford-Binet IQ at age 5½ and number of days of previous attendance at nursery school, with initial IQ constant, was .013. In subsequent follow-ups, the mean initial IQ on

the Minnesota Scale was compared with Stanford-Binet IQ at ages 5½, 6½, 8½, 10½, and 12½ for both preschool and non-preschool groups. Although both groups tended to improve in the later tests, the improvement did not favor the nursery group. In fact, many of the comparisons show a significant advantage in favor of the *non-nursery group.*

TABLE 9 *IQ Gain in Relation to Length of Nursery School Attendance*

(Data from Goodenough and Maurer, 18, pp. 169-171)

Group	One-year Retest		Two-year Retest		Three-year Retest	
	N	*Mean IQ Gain*	*N*	*Mean IQ Gain*	*N*	*Mean IQ Gain*
Preschool	84	4.6	51	6.2	13	5.8
Control	122	4.6	29	4.6	15	4.0

In explanation of the greater gains of the non-nursery group, the authors point out that *selective elimination* operated more markedly in the non-nursery sampling. Children dropped out of this group in greater numbers, and those who dropped out tended to be of lower IQ, than was the case in the nursery group. It was the children of the intellectually superior parents, in general, who tended to remain in the study in the non-nursery group. From this observation, the authors go on to suggest that, as children grow older, they approach more closely their "true" intellectual level and therefore come to resemble their parents more closely, the implication being that this resemblance is primarily a matter of heredity. It should be noted that, logically, the reported facts are equally consistent with an explanation in terms of the environmental effect of superior homes. Since the non-nursery children who remained in the study came from superior homes, development in such homes would stimulate a rise in the functions sampled by intelligence tests. The nursery group, according to the authors' own report, did not undergo so much selection in terms of home environment, and would therefore be less likely to improve. Whatever gains resulted from preschool attendance would thus have to counterbalance the greater gains resulting from home influences in the control group.

The fact that motivational differences were eliminated to a greater extent in the present study than in the Iowa investigations may also in part account for the divergent results of the two studies. As will be shown in a later section, such motivational differences may play a significant part in the observed effects of nursery school experience upon tested intelligence.

From the Institute of Child Welfare of the University of California,

Jones and Jorgensen (23) report data on a total of 54 nursery school children. Comparisons were made with control groups which had had no nursery school training but had participated in a similar program of serial mental tests. Socio-economic level and parental education were superior in both nursery and control groups. Follow-up studies included annual retests between the ages of 5 and 9. The tests administered were the California First Year Mental Scale up to 18 months of age, the California Preschool Schedule up to 5 years, the 1916 Stanford-Binet at ages 6 and 7, and the 1937 Stanford-Binet at ages 8 and 9.

For purposes of analysis, the total nursery group was subdivided into a number of smaller groups. First, 14 nursery children were matched with 14 non-nursery children in parental education. These two groups showed no significant difference in the "growth curves" of their test performance at any age. In another comparison, 11 nursery children were matched with 11 non-nursery children in mental test scores prior to nursery school attendance, as well as in the educational and occupational ratings of their parents. These two groups showed an increasing differentiation in test score with age, but no one of the differences in favor of the nursery group is statistically significant at any age. In this group, the parental occupational level proved to be slightly higher for the nursery group, the matching having been only approximately achieved. The authors point out that the actual discrepancy between the homes of the two groups may have been even greater than the occupational index indicated, since the parents who sent their children to nursery school at some financial sacrifice were probably superior. Thus in this study, the slight superiority in the home environment of the *nursery group* is offered as a possible explanation of their slight advantage in intellectual development.

A third group of 29 nursery children who had been given a number of different intelligence tests were compared with six control groups. Of the latter, two were matched with the nursery group on the basis of initial IQ, two on the basis of terminal IQ (i.e., IQ after nursery school attendance), and two on the basis of IQ's at ages 8 and 9. No significant difference between control and nursery groups was found in any of these comparisons, the various control groups differing more among each other than from the nursery group. For example, a group which had the same IQ as the nursery group at age 9 did not differ significantly from the nursery group in the testing prior to nursery school attendance. Or, conversely, a group matched with the nursery group in initial IQ did not differ significantly from it on subsequent tests.

Of particular interest is the analysis of length of nursery school attendance. In a group of 66 cases whose nursery school attendance varied from 50–99 to 450–499 days, the change in IQ on the California Preschool

Scale correlated .34 with length of attendance. The authors point out, however, that longer attendance was associated with greater number of testings in this study. As was shown in the preceding chapter, retesting will usually in itself raise scores. In the present study, the correlation between number of tests and IQ change was also .34. As might be expected from these data, the partial correlation between length of nursery school attendance and IQ change, when number of tests was held constant, proved to be only .05. This finding was corroborated in two other groups from the California growth study, consisting of 68 and 87 children. The corresponding partial correlations in these two groups were —.06 and .03, respectively.

From this brief examination of typical results it would seem that nursery school attendance may have a slight effect upon the test performance of most groups of children. The fact that so many of the groups studied come from superior home environments would tend to obscure the influence of the nursery school and may account for the lack of difference found in some investigations. Most of the nursery schools in which these investigations were conducted are connected with universities and tend, on the whole, to draw children from relatively superior families. In all but a very few studies the initial average IQ of the children was 110 or higher. If the children are already in an environment favorable for the development of those functions sampled by the intelligence tests, it is difficult to bring about additional improvements in IQ by special influences. A few of the investigators have recognized this factor, and there is some evidence that the effect of nursery school attendance may be greater with children from lower socio-economic levels (cf. 35, 39).

That the differences observed are often insignificant may result, too, from the very small groups employed in many of the studies. To find that a difference is insignificant under these conditions does not disprove the existence of a real difference—it merely fails to prove it. Such a finding certainly does suggest, however, that if there is a real effect of preschool attendance upon intelligence test performance, it must be slight. It has also been quite conclusively shown that, whatever the effect, it is manifested early, and that longer preschool attendance has little or no further influence upon the IQ. The interpretation of these findings will be postponed until the concluding section of the present chapter, since there are a number of methodological problems which must first be considered.

THE EFFECTS OF SCHOOLING FROM ELEMENTARY SCHOOL THROUGH COLLEGE

A few studies have dealt with the effects of schooling at the elementary, high school, or college level upon intelligence test performance. Some of these investigations report only retest results on a *single group*—a fact which precludes a clear-cut interpretation of their results. For example, retests of college students with the American Council Psychological Examination after 1 to 4 years of college work generally show a considerable rise in mean score (cf., e.g., 4, 19). These gains may be wholly or in part the result of simple repetition of the test or of retesting with a parallel form. They may also reflect in part the general improvement which the group would have made within a year even without attending college. What the net effect of the college experience was in producing the observed gains cannot be determined solely on the basis of the given results.

In a study (53) on children enrolled in *three superior private schools* in New York City, the results are also difficult to interpret for somewhat different reasons. The Stanford-Binet records of approximately 3000 children, accumulated over a period of 20 years, were examined. Among these records were over 1100 retests given after an interval of 2½ years or more, during which the child had attended the particular school. A significant mean gain in retest IQ was found in one of the three schools, but not in the other two. Further analysis of the scores from the former school showed that the maximum gain occurred within a short period of residence, later gains being negligible. The investigator offers no conclusive explanation of these findings, but suggests the possibility that "subtle and unidentifiable selective factors" may have operated in the one school to produce the gains. The slight mean gains found in the other two schools are no greater than is usually found in the repetition of a test.

In connection with the Iowa project cited in the preceding section, Wellman (57) compared a group of 269 children attending the *university elementary school* with a group of 47 children attending *other schools*. The two groups were equated in age and IQ at the end of the preschool period. The average age at this time was 5½ years, and the mean IQ's were 120.5 and 121.0, for the university school and other school groups, respectively. The former group gained an average of 5.6 IQ points after nearly 4 years' attendance at the university ele-

mentary school, while the other group gained only 1.2 points during the same period.[8]

An interesting comparison was made in an investigation (66) on rural children enrolled in the first three grades of *consolidated* and *one-room schools* in the same rural area. Stanford-Binet IQ's obtained in the fall and spring for two years showed significant gains during the school sessions on the part of the consolidated school children, but only a slight change or a loss in the one-room schools. This difference was not related to family background or to home environment, but is attributed by the author to the superior educational facilities afforded by the consolidated schools.

Another approach is illustrated by studies on the relationship between *amount of education* and intelligence test score. That a considerable relationship exists has long been a familiar fact. During World War I, when intelligence testing was still in its infancy, a correspondence between amount of schooling and intelligence test score was clearly demonstrated. Thus in a sampling of 48,102 recruits, the correlation between Alpha score and extent of schooling was .74 (cf. 67). A similar relationship between AGCT score and extent of education was found in World War II.[9] The establishment of such a relationship, however, does not in itself enable us to choose between the two alternative explanations, viz., (1) education raises the intellectual level, or (2) the brighter individuals are more likely to "survive" the increasingly stringent selection of the successive educational levels. That the duration of any one individual's education is not entirely dependent upon his ability is fairly obvious. Financial resources, family tradition and attitudes, educational facilities in different localities, and a number of other non-intellectual factors can readily be cited.

An interesting effort to secure data bearing more directly upon this question is reported by Lorge (33). In 1921–22, 863 boys, constituting a representative sampling of the public school population of New York City, were tested in the eighth grade with a number of psychological tests. Included among these tests were a reading and an

[8] The gain of the university school children has a critical ratio of 2.54 and is thus moderately significant, although falling short of the conventional criterion of 3.0. The control group gain is quite insignificant, having a critical ratio of .41.

[9] Cf. 7 for the mean AGCT scores of men with different amounts of schooling. In a sampling of 4330 men, a correlation of .73 was found between AGCT score and highest grade completed in school (cf. 47, p. 765).

arithmetic test which together yield a composite score reported to be essentially equivalent to the scores on most intelligence tests. Twenty years later, 131 of the original subjects, shown to be a representative sample of the total group in terms of original test means and SD's, were given two group intelligence tests. Typical results on one of these tests are reproduced in Table 10.

TABLE 10 *A Twenty-Year Follow-Up on the Effects of Schooling upon Intelligence Test Performance*

(From Lorge, 33, p. 487)

Initial Intelligence Test Score in 1921:	49–58		59–68		69–78	
Highest School Grade Completed	Otis Score	N	Otis Score	N	Otis Score	N
8	14.0	4	22.0	4	20.7	9
9	19.0	1	19.5	2	14.5	2
10	24.0	1	22.0	4	25.1	9
11–12	21.0	1	26.0	1	31.7	3
13–14			22.0	1	26.0	3
15–16			34.0	1	27.0	1
17 or more					38.0	3

Initial Intelligence Test Score in 1921:	79–88		89–98		99–114	
Highest School Grade Completed	Otis Score	N	Otis Score	N	Otis Score	N
8	26.4	5	39.0	2	33.0	1
9	31.1	8	38.0	2	29.0	1
10	28.5	8	37.0	4	46.5	2
11–12	31.0	9	41.0	3	34.0	1
13–14	34.7	4	41.7	4	37.5	2
15–16	39.5	6	53.5	2	50.8	5
17 or more	46.0	5	54.5	6	43.0	1

It will be noted that individuals who fell within a single class-interval in the original test, but who completed varying amounts of education in the interim, differ considerably in their performance on the 20-year retest. The mean Otis scores within each initial category show a fairly consistent rise as education rises. For example, among the 30 men whose 1921 scores fell between 69 and 78 (third column in Table 10), the 9 who completed only the eighth grade obtained a mean Otis

score of only 20.7 in 1941. The 3 who had taken graduate training beyond college averaged 38.0.

If we consider the last four columns in Table 10, in each of which a full complement of educational levels is represented, and if we combine extreme groups in order to deal with somewhat larger samples, we obtain the summary data shown in Table 11. According to

TABLE 11 *A Comparison of Extreme Groups from Table 10*

Initial Intelligence Test Score in 1921:	69–78		79–88		89–98		99–114	
Highest School Grade Completed	Otis Score	N	Otis Score	N	Otis Score	N	Otis Score	N
8–10	22.1	20	29.0	21	37.8	8	38.8	4
15 or more	35.3	4	42.5	11	54.3	8	49.5	6

these figures, a difference of about 7 or 8 years in schooling led to a mean difference of from 10.7 to 16.5 points in intelligence test score. As a means of gauging the magnitude of this difference, we may compare the groups which received the same amount of education, but differed in initial intelligence test scores, i.e., reading *across* Table 10. For example, among the subjects with only eighth grade education, the mean Otis scores obtained by groups which differed in their initial test scores range from 14 to 39. Among those with 11 or 12 years of schooling, the means range from 21 to 41. Within the highest educational level, with one or more years of academic work beyond college, the range is from 38 to 54.5. Thus the differences between columns in Table 10 appear to be about as large as those between rows. In other words the inter-group differences in adult scores were approximately as large when education varied as when initial score varied

In evaluating such a finding, a number of points should be borne in mind. On the one hand, the subjects upon whom these comparisons are based were not *exactly* equated in initial performance (cf. 14). Each of the vertical categories in Tables 10 and 11 covers a class-interval of 10 points in initial score. From this fact, it might be argued that those boys falling near the top of this initial range in any one category were the very ones who continued their education longer and received higher scores 20 years later. The conclusiveness of the

demonstrated effect of schooling is further limited by the small number of cases involved in some of the comparisons and by the fact that both initial and final tests inevitably fell short of perfect reliability. The chance errors thus introduced undoubtedly affect the amount—and possibly in some cases even the direction—of the obtained differences.

These considerations would suggest that the apparent effect of schooling in the above study may be slightly overestimated. On the other hand, it should be noted that all the subjects in this study had the first eight years of schooling in common. They had, in fact, attended the same classes during their first eight years. This "constant," added to each subject's education, would certainly make the individuals more nearly alike than they would have been had some of them received less than eight years of education. If the range of education within the group had been wider—say, from the third grade to college graduation—there is no doubt that the net effect of schooling on intelligence test score would have been greater. In fact, it is not unlikely that on most current intelligence tests the effect of the first eight years of schooling is greater than that of subsequent education. Beyond elementary school, the content of instruction is less standardized and uniform, and therefore less likely to be sampled in the construction of intelligence tests. Thus the present study demonstrates that extent of education can influence intelligence test performance. But it would be unwarranted to generalize regarding the extent of such effect beyond the specific conditions of this study.

Relevant data are also to be found in a comparison of the intelligence level of soldiers in World Wars I and II (55). A group of 768 enlisted men, representative of the entire population of white enlisted soldiers in World War II, were given both the AGCT and a revision of the Army Alpha. The distribution of this group on the AGCT paralleled very closely that of the entire army. On the Alpha their median score was 104, in contrast to a median of 62 obtained in World War I. The magnitude of this difference can be more clearly envisaged when we consider that the median of the World War II sampling corresponds to the 83rd percentile of World War I. In other words, 83% of the World War I group fell below the median score of the World War II sample. A number of factors may help to account for this marked improvement in intelligence level over the twenty-five years. Among them may be mentioned the later

group's greater experience in taking tests in school, in industry, and in the army itself. The possible influence of better physical condition, as a result of improvements in public health and in nutrition, should also be considered. The major factor, however, appears to be the higher educational level of the population, together with probable improvements in the quality of instruction, length of school term, and the like. In the World War II sample, the average education was 10.0 years, i.e., two years of high school. The comparable World War I average was 8.0, or elementary school graduation.

METHODOLOGICAL PROBLEMS

The methodological problems characteristic of studies on the effects of schooling arise largely from two necessary conditions of such investigations, viz., longitudinal observation and the comparison of matched groups. We shall consider some of the most persistent of these problems under seven major headings. Some are concerned with the choice of subjects, others with the measuring instruments or the particular conditions to be controlled in the course of the observations. A few deal with broader questions of the general plan or experimental design of this type of investigation. All these points have been cited and elaborated in the course of the controversy regarding the effect of schooling on the IQ.[10] The present section is not, however, intended as a summary of the criticisms which the various workers in this field have directed against each other's research—the list would need to be much larger in that case! Our present concern is with the more general methodological problems which every investigator in this field must face, rather than with the minutiae of the shortcomings of specific studies.

Sampling Problems. Considerations of sampling, or the choice of subjects to be investigated, enter into "schooling" studies in several ways. First, because this type of investigation is generally based on longitudinal, "follow-up" observations, it is likely to be automatically restricted to a *selected sample* of the general population. Stability of residence and continued cooperation of parents would, for example, be necessary conditions for the inclusion of children in a follow-up study of several years' duration. A group which is "selected"

[10] See especially 15, 16 22, 36, 37, 38, 51, 60, 63.

in terms of these conditions may in turn show other characteristics related to cultural level of the home, parent-child relationships, and the like. For these reasons, it is likely that the samplings employed in longitudinal studies tend to be somewhat superior to the general population. The reverse may be true in the case of institutional samples, such as orphanage children. In this situation, the superior members may, for example, be more often removed for adoption. The enduring sample would thus represent an inferior selection. In either case, generalizations from a longitudinal sampling to the total population must be made with considerable caution and with due regard for the selective factors which may have operated in the particular situation.

A second source of sampling difficulty pertains to the *matching* of experimental and control groups. Ideally, matched groups should be set up in advance by the experimenter, from the same population. In testing the effects of nursery school attendance, for instance, the experimenter would pair off children *in advance* on the basis of matching characteristics, and would then assign one member of each pair to the nursery group and the other member to the control group. The choice within each pair would be purely random.

In actual practice, investigations of schooling have had to resort to *a posteriori* matching. Certain children within a community are entered in nursery school on the basis of their parents' decision. Such a decision may itself reflect characteristics which distinguish these parents, their homes, or their children from others in the community. The investigator now steps in and tries to find other children in the community who "match" these nursery children in what he considers to be important characteristics for his study. The difficulty lies in the possibility that one or more characteristics whose relevance to the problem at hand may have been overlooked will now be allowed to vary between the two groups. If the assignment to nursery and non-nursery groups had been made in advance by the experimenter, these uncontrolled characteristics would probably vary at random in the two groups and no serious error would have been introduced in the results. But if special factors, such as parents' decision to register their child in nursery school, determine the placement of the child in experimental or control group, then the uncontrolled characteristics may vary systematically, piling up an excess of one type of child in only *one* of the groups.

If, for example, children from the more "progressive" or "enlightened" homes are sent to nursery schools, then the systematic difference in home atmosphere in favor of the nursery group might in time lead to superior development of this group, in contrast to the control group. Or it might happen that children who are inclined to be shy are more often sent to nursery school to enable them to overcome this difficulty. In such a case, the child's shyness might handicap him on his initial intelligence test and lead to an apparent gain on a later test, when the shyness in the unfamiliar situation had decreased. These examples are given merely to point up the dangers of *a posteriori* matching. No investigator can foresee or even subsequently identify all relevant characteristics in which his control and experimental groups should be equated. The random assignment of individuals to the two groups in advance is therefore an important safeguard against systematic differences in unmatched characteristics between control and experimental groups.

In any comparison between matched groups, it is of course essential that the groups be equivalent at the time when the comparison is made. Groups which were originally matched closely may become quite unlike through the *selective dropping out* of individuals, a selection which may operate differently in experimental and control groups. Similarly, it would obviously be misleading to compare the average initial IQ of 100 children with the average IQ of 13 of these children who have remained in the study four years later. The only significant comparison in such a case would be that between the initial scores of these 13 children and their own final scores.

Finally, the *size of sampling* needed in this type of investigation should be considered. Many of the studies have been conducted on very small samples. Even when several hundred subjects are included in the investigation, specific crucial comparisons have often been made between sub-groups of less than 50 cases. If the effects of schooling on intelligence test performance were very large and clear-cut, small samples would suffice to demonstrate the relationships under consideration. But the effects of schooling constitute a small part of all the influences which make for similarities and differences among individuals or groups. When effects are relatively slight, they may readily be obscured by chance factors in a small sample and only insignificant differences will be obtained.

Statistical Regression. As a general statistical concept pertaining

to correlated measures, regression has long been familiar. Its particular application to the type of investigation under consideration (cf. 36, 52) is based upon: (a) certain *sampling problems* arising from the use of matched groups, which were discussed in the preceding section, and (b) the fact that *test reliability* is always short of perfect, i.e., every test score contains some "error of measurement." Upon retesting, regression may artifically produce two distinct effects, one pertaining to the relative position of individuals within the group, and the other to the relative standing of the two matched groups. We shall begin by considering the first of these two effects, since it is the simpler of the two.

Statistical regression simply means that extreme scores on an imperfect measure of any characteristic tend to "regress" or move toward the mean upon retesting. Such an effect occurs when two different tests of the same characteristic are given, as well as upon the repetition of a single test. For example, if a test for speed of tapping is administered to 100 subjects on Monday and again on Wednesday of the same week, a tendency will be found for those who scored far above the group average on Monday to fall closer to the average on Wednesday, and for the Monday low scorers to rise toward the average on Wednesday. Similarly, if a group of children are tested with the Stanford-Binet and the Merrill-Palmer Scales, those receiving high Binet IQ's will tend to drop on the Merrill-Palmer and those with low Binet IQ's will, in general, show a gain on the Merrill-Palmer.[11] It should be noted that the regression effect does not depend upon the sequence in which the tests are administered. It simply occurs in any comparison between the scores of the same group of individuals on two imperfectly correlated measures. For example, if we select the children with the highest Merrill-Palmer IQ's in the above illustration, we shall find on the whole that their Binet IQ's are not as high as their Merrill-Palmer IQ's; those with low Merrill-Palmer IQ's will tend to do better on the Binet than on the Merrill-Palmer. Similarly, in the tapping illustration, the Wednesday high scorers will tend to have performed more poorly on Monday, and the Wednes-

[11] This example is taken from actual results obtained in one of the Iowa nursery school studies (cf. 59, p. 98), in which IQ's on the Stanford-Binet (or Kuhlmann-Binet for younger children) were compared with Merrill-Palmer IQ's. The regression effect refers, of course, to the individual's *relative position* in the group, not to absolute differences resulting from test standardization, practice, and th like.

day low scorers will have done better on Monday. Thus in any comparison between two measures which are not perfectly correlated, regression occurs in *both directions*, i.e., it is a reversible effect.

Such a regression effect results entirely from the "error of measurement" in the scores. Thus, some individuals who received high scores on the first test did so in part because an error of measurement raised their score *on that particular occasion*. Since such errors of measurement are uncorrelated on two testings, this person will probably score lower upon retesting, i.e., by chance the error will be unlikely to occur in the same direction on both occasions. By the same token, some of the individuals receiving low initial scores did so because chance factors lowered their score on that particular occasion. To the extent that this was true, these individuals will tend to gain on a retest.

It should be noted parenthetically that the error of measurement to which we refer need not be an "error" in the popular sense. It is an error only in so far as the two tests are attempts to measure the same behavior and would therefore be expected to yield identical scores. Any factor specific to one of the tests and not entering into the other would tend to make the two scores unlike and would constitute an "error" for the present purpose. In the above illustrations, such an error would reflect the different fortuitous influences which might raise or lower performance on any day on the tapping test. It would also include the factors specific to such tests as the Stanford-Binet and the Merrill-Palmer—factors which differ from test to test despite the fact that both tests are designed to measure intelligence. In its broader applications, then, regression occurs between any two measures whose correlation is less than 1.00 and which therefore include *specific factors* differing from one measure to the other.

Not all changes in score, of course, are the result of regression. With a highly reliable measuring instrument, the error of measurement will be slight and the scores will reflect more accurately individual differences in the ability being measured. It follows under these conditions that changes in score from test to retest will depend more largely upon actual gains and losses in the ability under consideration. With less reliable measures, however, the error of measurement constitutes a relatively large part of the score, and regression effects will be greater. In the testing of children of preschool ages,

the reliability of intelligence tests is relatively low.[12] For this reason, the regression effect becomes an important problem in the evaluation of data obtained on preschool groups.

An observation repeatedly reported in a number of nursery school studies, both with superior nursery groups and with orphanage children, is that the brighter individuals in the group tend to gain least or even to lose following nursery school attendance, while the duller members make the largest gains. This has been interpreted by some as a "leveling" effect of the nursery school experience. It has been argued that the "stimulating value" of each specific environment tends to make individuals approximate a particular intelligence level. Individuals above this critical point in their initial IQ will gain nothing and may even be "pulled down" to the performance level corresponding to the environment in which they have been placed; those below this critical level, on the other hand, will be raised by the environmental stimulation.

The correctness of this explanation can always be checked by comparing the *total group variability* before and after the interpolated experience (cf. 36). If, for example, nursery school attendance really has a leveling influence upon the IQ, then the range of individual IQ's should decrease from the initial to the final test. Such a decrease in variability should likewise be discernible in a drop in the SD of the group. In the nursery school studies it has been demonstrated, however, that *individual differences do not decrease significantly during preschool attendance*. The number of cases at different IQ levels tends to be the same before and after such an experience, although *different individuals* fall into each IQ level on the initial and final tests. Thus what is actually occurring is that individuals are merely trading places upon the retest, rather than undergoing a leveling of ability.

The distinction between leveling and regression is illustrated schematically in Figure 46. For simplicity of discussion we shall assume that both distributions have identical means of 100 as well as identical variabilities. Part A of this figure shows the effect of regression upon 10 hypothetical individuals selected because each received an IQ of exactly 120 on Test 1. It will be noted that, owing to "chance errors" and specific factors in the scores on Tests 1 and 2, these 10 persons "fan out" on Test 2. The average of the 10 IQ's, however, is nearer to the distribution mean on Test 2 than

[12] The reasons for such low reliability will be discussed in a subsequent section on the instability of early IQ's.

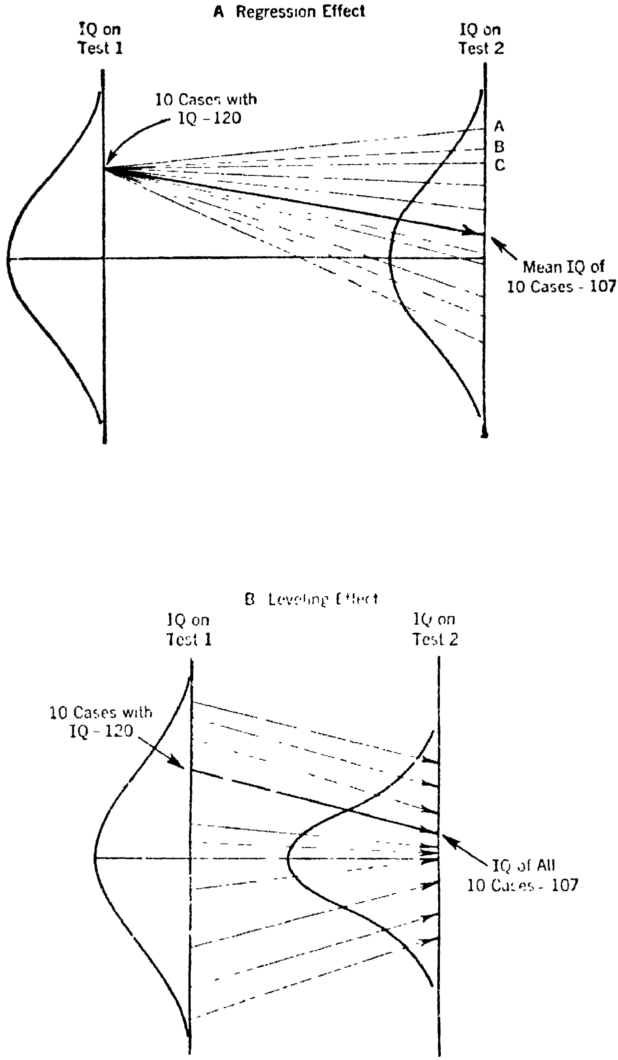


Fig. 46. Contrast between Regression and Leveling Effects.

it was on Test 1. Thus the mean of these 10 cases was 120 on Test 1, but is only 107 on Test 2. It is only in this sense that the 10 individuals have regressed toward the distribution mean. At the same time, it should be noted that certain *individuals* who score 120 on Test 1 actually fall farther from the distribution mean on Test 2 than on Test 1. In the diagram, this is true of persons A, B, and C. The reader should visualize a similar "fanning out" of scores throughout the distribution of Test 1. This means, first, that individuals receiving any one score on Test 1 are likely to spread over several scores on Test 2. Secondly, the average of these Test 2 IQ's will not fall as far from the Test 2 mean as the Test 1 IQ of these same individuals diverged from the Test 1 mean.

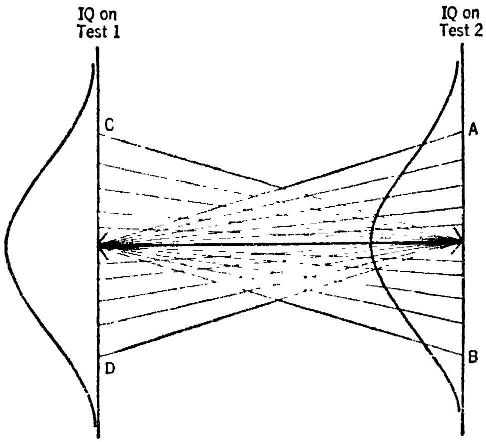


Fig. 47. Reversibility of the Regression Effect.

Leveling effect is illustrated in Part B of Figure 46. Ten hypothetical persons having different IQ's on Test 1 are shown to have moved closer to the mean on Test 2. If only leveling operates, this movement toward the mean would occur in each individual. Moreover, if we assume for clarity of illustration that no regression occurs at all in this comparison, but only leveling, then 10 persons with identical IQ's of 120 on Test 1 will also have identical IQ's on Test 2, although they will be closer to the mean on the latter test. In the illustration given in Figure 46, the Test 2 IQ's of these 10 persons are all 107. It is apparent that, in true leveling, the variability of the group should drop from Test 1 to Test 2. Such a drop does not occur when only regression operates.

An illustration of the "reversibility" of the concept of statistical re-

gression is given in Figure 47. Beginning with 10 hypothetical individuals whose IQ's are all exactly at the mean of Test 1, we see that on Test 2 their IQ's spread over a considerable range. If, on the other hand, we begin with 10 individuals whose IQ's fall at the mean of Test 2, the identical effect is found, in the reverse direction, on Test 1. Since, unlike leveling, the regression effect occurs equally in both directions, it is obvious that variability cannot decrease.[13] The fact that individuals only trade places in the regression effect, without affecting the range of scores, may be vividly illustrated if we observe the positions of individuals A, B, C, and D in Figure 47. It will be noted that A and B move from the distribution mean on Test 1 to a high and a low IQ, respectively, on Test 2. Individuals C and D, on the other hand, who had a high and a low IQ, respectively, on Test 1, fall at the distribution mean on Test 2. Whenever any one individual shifts *toward* the mean from Test 1 to Test 2, he obviously shifts *away* from the mean by an equal amount from Test 2 to Test 1.

So far, we have discussed regression and leveling effects only with reference to *individuals.* Under certain conditions, however, regression may also produce a difference between the means of two initially matched *groups.* This will occur if the populations from which the samplings are drawn differ in the characteristic in terms of which the groups are matched. Regression of scores on a second test occurs *toward the mean of the population from which the cases are selected.* If these means differ appreciably, then two samples of these populations, which were deliberately matched in initial score, will diverge on a retest in the direction of their respective population means.

For example, men and women as a whole will differ in performance upon a test of strength of grip.[14] The male population will have an appreciably higher mean than the female population in this characteristic. If, now, we select a sample of men and women who are matched on a test of strength of grip, we will have to choose men from the lower end of the total male distribution in this characteristic, and women from the upper portion of the female distribution. A retest of strength of grip administered to these matched samples will

[13] In *predicting* scores from a regression equation, the variability does decrease. Thus if the measured and the predicted scores correlate .75, the SD of the predicted standard scores will be only three-fourths as large as that of the actual standard scores. In the present situation, however, we are dealing throughout with *actual* scores.

[14] This illustration, as well as the hypothetical data which follow, are taken from R. L. Thorndike (52, p. 91).

show that the women have regressed downward toward their population mean, and the men have regressed upward toward theirs. This assumes, of course, that each strength-of-grip test falls short of perfect reliability.

As long as the category in terms of which the population is defined (in this case, sex) is itself related to the characteristic in which the subjects are matched (i.e., strength of grip), then the means on a retest will regress toward the respective population means. This follows from the fact that whenever we choose individuals from the upper end of the total distribution, we are capitalizing on errors of measurement, i.e., on whatever component in the scores of Test 1 is uncorrelated with the scores of Test 2. We tend to choose individuals who have scored high not only in the components shared by Tests 1 and 2, but also in those specific to Test 1. The scores of the majority of such individuals are thus likely to drop on Test 2. The reverse will be true if we choose a sampling of individuals from the lower end of the Test 1 distribution.

A clear illustration of this regression effect is furnished by R. L. Thorndike (52, p. 91), using artificial data derived from dice throws. Such data have the advantage of demonstrating the regression effect which follows as a mathematical necessity from the given conditions, without the confusing interference of other unknown variables which might operate with real subjects. The relevant data for this illustration will be found in Table 12. Scores corresponding to initial test and retest were found for 132 "men" and 132 "women," constituting the two populations.[15] The mean difference between these two populations on the initial test proved to be 4.7. From these two distributions, matched samples of 64 men and 64 women were selected. The means of these two matched samples on the initial test were, of course, identical, each being 20.3. When the retest scores of these 64 men and 64 women were examined, however, the means of the two groups were 21.0 and 19.2, respectively. Thus the two samples which had been matched on the initial test regressed toward their respective population means on the retest.

[15] For each individual, 7 dice were thrown. The score on the initial test was the number of spots showing on dice 1, 2, 3, 4, and 5, and the score on the retest the number of spots showing on dice 1, 2, 3, 6, and 7. In determining the "men's" scores, the same dice-throwing procedures were followed, and a constant value of 5 was then added to each score.

TABLE 12 *The Effect of Regression upon Matched Groups*

(Data from R. L. Thorndike, 52, pp. 89-90)

	Population: Initial Test Means	*Matched Samples: Initial Test Means*	*Matched Samples: Retest Means*
Men	22.2	20.3	21.0
Women	17.5	20.3	19.2
Difference	4.7	0.0	1.8

The similarity between the pattern of this hypothetical situation and that represented by the nursery school studies is clearly apparent. If preschool children as a whole (i.e., the preschool population) tend to come from superior homes and to have higher IQ's than non-preschool children (for a variety of reasons unrelated to nursery school attendance), then upon retesting, the matched preschool group will tend to gain in IQ and the non-preschool to drop or remain unchanged. Each will thus have regressed toward its own population mean. In such a case, the matching of the two groups in initial IQ must have been accomplished by the inclusion of children from the lower end of the preschool population and the upper end of the non-preschool population, as in the hypothetical male and female populations of the above illustration. The extent of this regression effect depends upon the amount of difference in initial test between the two populations, as well as upon the test-retest reliability. The relatively low reliability of intelligence test scores for young children, coupled with the practice of matching samplings *a posteriori*, would thus make regression a serious problem in the preschool studies.

The Role of the Examiner. In any retest study, the "*personal equation*" of individual examiners must be taken into account. In longitudinal studies with the Stanford-Binet on school children, for example, the IQ's obtained by different examiners on the same group of children have been found to vary considerably (cf., e.g., 10). Some examiners give consistently higher and some consistently lower IQ's. Unlike group tests, most individual scales permit sufficient latitude in both administration and scoring to make the role of the particular examiner a considerable one. In the testing of young children, the part played by the individual examiner is even more prominent. Whenever long-range studies necessitate shifts in personnel, it is

therefore essential to check for any systematic variations in the results of different examiners. A gain or loss of a few points from initial to final IQ might otherwise be attributed to an actual change in performance, when in reality it resulted from the use of different examiners in the administration of the two tests. A lack of difference between initial and final IQ might likewise be a spurious result, if a more lenient initial examiner happened to be balanced against a real gain in the subsequent test.

An even more important precaution is to rule out any possible effects of the examiner's *"mental set."* This can be successfully accomplished only when the examiners are ignorant of the child's previous test records and of his experimental classification. The examiner should not, for example, know when he is testing a child in the nursery group and when he is testing one in the control group. Without such a safeguard it is very difficult to prevent unintentional bias from operating in either the administration or the scoring of the tests. This is, of course, the type of precaution regularly followed in any well-conducted laboratory study. But it has rarely if ever been thoroughly applied in studies on the effects of schooling, especially at the lower age levels.

An "expectation" that a particular child will do well or poorly may be established through the examiner's knowledge of the child's previous performance, or through the examiner's personal hypothesis regarding the probable outcome of the experiment. The halo effect in any situation calling for ratings or judgment is a familiar example of the influence of such expectations. But its operation in more objective testing situations has also been demonstrated. Goodenough (16), for example, cites a study on the errors made by school teachers in grading spelling papers. A correlation of .40 was found between the number and direction of such errors and the ratings for "personal attractiveness" given by the teachers to the same children. Children who were rated as more attractive by a particular teacher thus tended to gain by the clerical errors made in the scoring of their papers; those rated as less attractive tended to lose. These errors occurred under conditions in which the teachers were endeavoring to avoid making *any* errors!

Emotional and Motivational Changes. Part of any demonstrated improvement in test performance following nursery school attendance

may result from better emotional and motivational adjustment of the child to the testing situation. In such a case it would be necessary to consider the extent to which such improved attitudes toward adult-determined tasks might exert a general effect upon the child's learning and intellectual development.

That frustration and other emotional experiences can significantly affect the level of performance on the Stanford-Binet or other similar tasks is suggested by evidence on both preschool and school children (30, 31). Especially important in this connection is the characteristic resistance or negativism often encountered in young children. Such behavior serves as a constant error, always lowering and never raising the score. The extent to which this error can influence test scores was demonstrated in a study on approximately 100 three-year-old children (44). Those tests which the child had refused to perform—but not those that he had failed—were repeated on successive days until they were definitely passed or failed. The effect of this procedure was to raise the Kuhlmann-Binet [16] IQ in over two-thirds of the cases. Among these, 18 children gained from 15 to 24 points, and 7 gained as much as 25 to 35. There is a strong probability that nursery school attendance may reduce negativism, especially in a psychological test which has much in common with the nursery school situation.

Comparability of Measuring Instruments. The use of *different intelligence scales*, which may not be comparable, *at different age levels* is obviously a complicating factor in any longitudinal study. To be sure, such a procedure does not in itself affect the comparison of experimental and control groups, since both groups take the same test at the same time. It should be noted, however, that groups matched on one test cannot be assumed to be matched for other tests which may sample somewhat different behavior functions. Hence it is obvious that the use of different intelligence scales for initial and final testing would make the results confusing and difficult to interpret.

The use of different scales *by different investigators* may also account in part for the apparent inconsistency of their results. It has been pointed out, for example, that certain tests, such as the Merrill-Palmer Scale, are very similar in content to typical nursery school

[16] Similar results were obtained with the Merrill-Palmer Scale, although the gains were considerably reduced when a correction for refusals had been made in the original IQ's. Even these corrected scores, however, were raised from 1 to 14 IQ points in over one-fourth of the cases, upon completion of the "refusal" tests.

activities. Improvement on such tests following nursery school attendance is thus likely to be highly specific to the test functions and not diagnostic of other behavior.

What is perhaps less obvious—although equally important—is that the *same scale* may not yield strictly comparable results *at different ages*. Many tests, including the Stanford-Binet, measure rather *different functions* in the early and later age levels. At the lower ages, such tests cover largely sensori-motor coordination, sensory discrimination, memory for objects, and similar simple behavior. As age level rises, increasing emphasis is placed upon verbal and other symbolic functions. In a study of many years' duration, misleading differences between control and experimental groups might appear simply because the groups have been equated closely in non-verbal but not in verbal functions. Thus groups which (for any reason) are more adept in verbal and other symbolic functions than they are in sensori-motor activities are likely to show a gain in IQ upon later retesting, regardless of any specially interpolated influences. To be sure, such a gain would also occur in the control group, provided that matching is adequate and extends to all relevant factors, such as home background. But if the subjects are matched only in initial IQ, as is often the case, a difference might appear between them in later tests, which has no relation to the interposed experiences. This difficulty cannot be completely avoided if testing begins at very early ages, when the repertory of verbal behavior is still largely undeveloped.

A further difficulty in the use of certain scales, such as the Merrill-Palmer, in longitudinal studies is that the *meaning of a score* varies at different age levels, owing to the statistical characteristics of the norms. For example, an IQ of 114 at one age may indicate the same degree of superiority as an IQ of 141 at a later age.[17] Even in such carefully constructed tests as the 1937 Stanford-Binet, spurious IQ changes may occur at different age levels (cf. 17). The extent of individual differences in IQ on this scale is not constant at all ages. Studies by several independent investigators have confirmed the observation originally made by the authors of the test that IQ variability is highest at age 2½ or 3, lowest at about age 6, and again reaches a

[17] This follows from the fact that the SD of the mental ages on this test does not increase with chronological age in such a way as to yield constant IQ's at successive ages. Such an increase in variability is a statistical prerequisite for the obtaining of IQ's which are comparable in meaning at different ages

high peak at age 12.[18] These differences in variability are large enough to produce age changes of 15 to 20 points in the IQ of children who are far above or far below the group mean (2 to 3 SD's away from the mean). Changes of 8 to 12 points in the *mean IQ* of groups composed largely of superior or of retarded children can likewise be brought about by these conditions.

When the subjects in the control and experimental groups are individually matched in age, such shifts in the meaning of an IQ do not constitute a serious difficulty, although they certainly introduce confusion and awkwardness into the interpretation of results. If, on the other hand, the groups are only roughly equated in age, then serious error may result in the comparative evaluation of IQ changes in the experimental and control groups. Moreover, any comparison between samplings from different populations, varying in home background, parental intelligence, etc., might be completely vitiated by these characteristics of the measuring instrument. Any gap in performance level initially existing between such groups may be spuriously enlarged (or reduced) at different age levels.

Care should also be taken to insure that the groups to be compared are given *an equal number of intelligence tests*. Otherwise it may often happen that children in nursery school, or those attending a superior school, receive more practice either on the same test employed in the experiment or on closely similar tests. It will be recalled that scores on the second or third administration of a test are not directly comparable to scores on its first administration.

Instability of Early IQ's. That IQ's obtained in infancy and early childhood are relatively unstable is not surprising in the light of much of the preceding discussion. In a follow-up of 91 children of superior socio-economic and intellectual level (2), scores on items from a variety of widely used infant tests, administered from 3 months to 5 years of age, showed little relation to Stanford-Binet IQ's at age 5. In another study (9), 138 children who had been tested between the ages of 2 and 6, during the standardization of the 1937 Stanford-Binet, were re-examined ten years later. When scores on the same form of the test were compared, the group initially tested at ages 2

[18] That these changes in variability with age may themselves have an environmental explanation—in terms of the increasing and decreasing uniformity of individual experiences between these ages—is beside the point. The fact remains that such changes occur in the absence of any experimentally interpolated factor such as special training introduced in any particular longitudinal study.

or 3 showed an average change of 13 IQ points; those children initially tested at ages 4 or 5 showed a mean change of 11 IQ points. The author's conclusion from these findings is typical of the current view of most psychologists: although "the Revised Stanford-Binet Scale is as good as or better than any other objective index in predicting the future intellectual functioning of a preschool child . . . an individual IQ obtained prior to the age of six years must be interpreted with discretion."

A number of factors are undoubtedly responsible for the low predictive value of IQ's obtained at the preschool ages. In the first place, the samples upon which the norms were established are generally not so representative as in the case of older groups, owing to the practical difficulty of gaining access to young children for testing purposes. The negativistic behavior characteristic of these ages, which may spuriously lower a child's score on any one testing, has already been discussed. It is also likely that "intelligence," heavily loaded as it is in our society with verbal ability, cannot be satisfactorily measured until the individual has attained a certain minimum of linguistic development. In a number of comparisons between the Stanford-Binet IQ's of 5-year-olds and their performance on various infant and preschool tests, L. D. Anderson (2) concluded that language development and linguistic items have the greatest predictive value. Comparison with later IQ's would probably show the predictive advantage of such verbal items to be even larger.

It should also be noted that, until they reach school age, most children have not been exposed to a sufficient body of uniform experience—later furnished by the relatively standardized school curriculum—to permit an adequate sampling of common intellectual tasks for testing purposes. That marked individual changes in IQ do not occur haphazardly, but may be related to experiential factors, is suggested by a further analysis by Bradway (9) of the 10-year changes in Stanford-Binet IQ cited above. Fifty subjects showing the largest test-retest changes were selected from the total group of 138 children. Detailed information on these cases was obtained through home interviews and visits, from which quantifiable data on thirteen home and familial characteristics were derived. A comparison of the 26 children showing IQ gains with the 24 who showed losses indicated that all of the factors studied were related to IQ changes.

Highly significant differences between the two groups were found, for example, in parental intelligence.[19]

Finally, mention may be made of the possibility that the relative "constancy of the IQ" observed at *later ages* is itself an inevitable mathematical consequence of the cumulative nature of behavior development.[20] The individual's behavior equipment at each age includes, in general, all his earlier behavior equipment, plus an increment of new acquisitions. Even if the annual increments bear no relation to each other, a growing consistency of behavior level would appear, simply because earlier acquisitions constitute an increasing proportion of total behavior as age increases. Predictions of IQ from 10 to 16 would thus be more "reliable" than from 3 to 9 because the scores at 10 include a larger proportion of what is present at 16, while scores at 3 include a smaller proportion of what is present at 9.

It should now be apparent that the "instability" of early IQ's depends only in part upon shortcomings of the measuring instrument. Some of the observed instability follows from the characteristics of behavior development itself. Predictive value of the IQ over periods of more than a year cannot be regarded as synonymous with test reliability in the accepted sense (cf. 49, 63). If genuine changes in performance level occur during such an interval, the scores on a highly reliable instrument will—and should—change. Body weight at the age of six months, for example, may correlate very low with body weight at age 40. Prediction from the former measure to the latter would be hazardous, and yet such measures may have been obtained with scales of nearly perfect reliability. Logically, a test may have high reliability—and validity—at a particular age level, despite the fact that it does not permit accurate long-range predictions. From a practical viewpoint, such a test would still have good predictive usefulness, i.e., predictions could be made from the specific behavior sample of the test to other behavior of the child at that particular age level.

[19] Whether parental intelligence operated as an environmental influence or through some unknown hereditary structural characteristic is, of course, not indicated by such a study. The data are cited here only to suggest that the "instability" of early IQ's may not be wholly a result of the unreliability of the measuring instrument, but may be definitely traceable to intervening environmental influences.

[20] Cf. J. E. Anderson's (1) concept of "overlap," to be discussed more fully in the following chapter.

The investigator who conducts a longitudinal study involving the testing of young children ought certainly to take every precaution to insure high test reliability. Thus adequacy of norms, good rapport, and the elimination of the effects of negativism are important prerequisites for unambiguous results. But the remaining "instability of early IQ's," resulting from such factors as insufficient verbal development, lack of highly uniform training, or the sheer paucity of the early behavior repertory, ought to be regarded more as an observed datum of behavior than as a weakness of the testing procedure.

Seasonal Variation. Among the results cited by Wellman (60) as evidence that the observed changes in IQ were actually attributable to nursery school attendance is the fact that the mean score rose regularly in the spring and fell or showed negligible change in the fall. The explanation advanced is that during the summer months, when the children did not attend nursery school, they were not exposed to the stimulation which brought about the spring gains in score. Seasonal changes in test performance have, however, been found in other investigations at the preschool level, as well as at later ages. Several studies (20, 21, 32) of non-nursery children of preschool age have shown that, in these groups too, a larger gain in score occurs over the winter than over the summer interval. It thus appears that attendance at nursery school was not the factor responsible for the differential changes during summer and winter months.

Seasonal differences in traditional activities, such as the shift from outdoor and gross sensori-motor activities during warm weather to indoor games and closer adult contacts during the winter months, may affect the child's "rapport" in the test situation. An optimum "warming-up period" in the type of functions sampled by intelligence tests may bring the child to his peak performance level at some time in the late fall or early winter. A month-by-month analysis of groups taking their initial and final tests in different months, with a six-month interval, showed the maximum gains in November, December, or January (20, 21). A considerable number of other seasonal factors, including holiday periods, nearness to vacation time, weather conditions, mounting ennui after a long period of similar activities, and the like, may also contribute to these results.

Whatever the causes, however, the fact that seasonal variations in test performance have been observed needs to be taken into consid-

eration in interpreting retest results. Changes in performance which normally occur at different times of the year cannot be attributed to the influence of specially interpolated experimental factors, such as nursery school attendance. Moreover, temporary changes resulting from reaction sets, warming-up, and similar influences ought to be distinguished from more permanent experiential effects.

IMPLICATIONS OF THE EFFECTS OF SCHOOLING UPON TEST PERFORMANCE

In the light of the above survey of methodological problems, it appears that studies on the effects of schooling upon intelligence test performance, under present conditions, are not well suited for an analysis of the heredity-environment question. At the nursery school level, where the majority of the studies have been conducted, no effect of preschool attendance upon IQ has been *conclusively* demonstrated. It is likely that, when various methodological difficulties are eliminated, a slight effect remains, but this may be the result of improved rapport which is highly specific to the test situation. To find that preschool attendance does not directly improve intellectual functions is certainly not surprising. Nursery school curricula were not designed for this purpose and have little direct bearing upon the verbal and other symbolic behavior functions which constitute so large a part of "intelligence." Their expected effectiveness in "raising the IQ" is further reduced by the fact that, in most of the studies, the children's home environments were already furnishing superior intellectual stimulation.

As for the few investigations on the contribution of subsequent schooling, they too represent a relatively ineffective approach to the study of environmental influences. The experimental design of such studies involves more similarity than difference in the experiential background of the contrasted groups of subjects. Among the different types of schools compared in most studies, the uniformities of instructional techniques and facilities seem to be more conspicuous than their diversities. Similarly, the compulsory education requirements in America furnish a common core of initial training upon which subsequent differences in amount of education are superimposed. The elementary school years may be of particular importance in the establishment of

work habits and in the development of the types of behavior which are most prominently sampled by intelligence tests. Similarities in home environment and in other out-of-school experiences add to the common core. When subjects have so large and so important a part of their environment in common, the observable effect of any environmental differences among them will be diluted. In summary, we may say, first, that studies on schooling have not furnished satisfactory, conclusive proof of large environmental effects. Secondly, such a finding is to be expected because of the experimental design of most of these studies. The studies on special educational programs set up for specific groups represent a more direct and better-controlled approach, but the data gathered by this method are still meager, though suggestive.

A further implication of the schooling studies which needs to be examined pertains to the relationship between "intelligence tests" and "intelligence." Much confusion has resulted and conflicting assertions have been made because of unrecognized assumptions regarding this relationship. One finds, for example, a tendency for writers on both sides of the controversy to "blame the tests," for opposite reasons. Thus on the one hand, appears the statement that if the IQ is shown to be inconstant and subject to modification, then the tests must be an unreliable or unsuitable measure of "intelligence." It has been argued (13), for example, that if intelligence tests prove to be susceptible to environmental changes, they must be heavily loaded with "experience factors" and ought to be revised.[21] Such statements obviously contain the tacit assumption that "intelligence" is not susceptible to environmental influences. Obviously, a procedure which sets out to reduce the contribution of a particular factor precludes the possibility of studying its influence.

On the other side, we find the proposal that since "intelligence" is susceptible to environmentally determined change, less emphasis should henceforth be placed upon mental testing in the schools, in order to avoid "a harmful and unwarranted label for a child" (50, p. 536). It has even been argued (cf. 11) that intelligence testing is undemocratic! To maintain that, because intellectual functions are susceptible to improvement by training, we should stop testing intelli-

[21] Actually, an effort is generally made to *reduce* "differential experience factors" in the selection of items for inclusion in intelligence tests. In fact, some writers have even suggested that, because of this practice, the tests are loaded in favor of heredity! (cf., e.g., 64).

gence, is equivalent to suggesting that we stop observing behavior. It also seems to identify schooling with the entire reactional biography of the child, ignoring influences outside the school situation. The value of psychological tests has been empirically demonstrated in many situations One need not believe in the hereditary fixity of the IQ in order to employ psychological tests. Such tests enable us to gauge what the individual can do *in his present state of development* To expect the tests to predict what the person will be capable of doing 20 years hence is to demand that the psychologist be a fortuneteller. Since we cannot foresee (except roughly) the experiences which any particular individual will have within 20 years, we obviously cannot predict his behavior very accurately over such an interval. This is particularly true, of course, if the test is given at an early age, when only a small fund of behavior has already accumulated.

Both of these extreme views illustrate a failure to recognize the proper relationship between the test and the behavior which is being tested. Actually, every intelligence test is only a *sample of behavior,* a small part of the type of behavior which we call intellectual. It follows that any observation made regarding test performance is *ipso facto* an observation regarding behavior. There is no sharp distinction between those factors which influence test performance and those which influence intelligence. The difference is one of *degree,* or of breadth of influence. Thus certain factors, such as negativism in test situations, may be narrowly limited in their area of operation, although it is doubtful that any influence is completely restricted to the test situation. The development of certain work methods and the acquisition of techniques for the symbolical manipulation of materials undoubtedly have a broader area of application. The various conditions which affect test performance probably fall into a continuum in reference to the breadth of their influence upon behavior.

Finally, mention may be made of the notion implicit in some discussions that "intelligence" is not definable in terms of observable behavior, but is a hidden, unexplorable entity or potentiality which remains immune to change while behavior may alter conspicuously. Such a concept of an entity which is not susceptible to observation and about which no statement can therefore be either proved or disproved has no place in science.[22]

[22] For a further critical evaluation of the concept of "potentiality," the reader is referred to Chapter 2

CHAPTER 9

Age Differences

A DISTINCTION HAS FREQUENTLY BEEN MADE between development through specific practice or training in a given activity and development through maturation or "growth" (cf. Ch. 4). Such a distinction does not imply a dichotomy between inherited and acquired behavior. Thus maturation is not regarded as independent of environmental stimulation of a general sort, nor is learning necessarily considered to be exclusively determined by environmental factors. When we speak of growth, we usually think of a definite sequence of developmental stages in the structural characteristics of the individual. As the child grows older, for example, his height increases, his bodily proportions are altered, and many other well-known physical modifications occur. Such changes take place regardless of the specific training which the individual may have had.

As structures become altered with age, so we may expect their functions to undergo change. With stronger muscles, the older child can learn to walk, climb stairs, sit up, and perform various other tasks much more readily than his younger brother. It is reasonable to expect that certain types of activity will in general appear at fairly definite stages, since they require a specific degree of structural development for their execution. Very intensive training at an earlier age may produce almost negligible effects when compared with the achievements of an older child with only a minimum of training.

Since such a large share of infant behavior consists in the acquisition of motor skills and sensori-motor coordinations, activities which are closely linked to structural factors, growth rather than practice seems to play the major part in early behavioral development. It is quite a different matter, however, to use the concept of growth to describe the intellectual and emotional development of the older child.

Such a concept has nevertheless been commonly employed in interpreting age changes in mental test performance, and the curves plotted to portray these changes have been labeled "mental growth curves." Such growth curves are difficult to interpret for many reasons and their use has led to much technical controversy.

It should become apparent in the course of the present chapter that the distinction between investigations of "training" and those of "mental growth" is a superficial one. It is only for convenience, therefore, that the former were discussed in the preceding chapters, while the latter have been reserved for the present one. The data on both topics should be considered as a whole. A few studies, in fact, are difficult to classify into one or the other category; this is especially true of experiments on very young children, such as those reported in Chapters 5 and 6. For the purposes of the present discussion, however, we may regard studies of "growth" as those in which mental test progress at successive ages is observed and charted, with no attempt to alter the normal course of development.

THE GROWTH CURVE

Growth curves were first plotted to show the development of physical traits, such as height, weight, bodily proportions as indicated by various indices, and the like. An example of such a curve, showing the changes in height in groups of tall, average, and short girls between the ages of 5 and 17, is given in Figure 48. As a descriptive technique for portraying more vividly the course of development of structural characteristics, the growth curve has proved serviceable and intelligible. The physical data are relatively easy to interpret and unambiguous. By analogy, however, attempts have been made to plot curves of "mental growth," a procedure which has brought additional confusion into an already difficult problem. At best these curves are only a descriptive summary of changes produced by a multiplicity of factors. By lumping all such factors together and giving them a semblance of systematic growth, the main issues are often obscured.

We shall first examine some of the principal *methodological problems* met in the measurement of "mental growth." A consideration of such problems goes far toward explaining the discrepancies and disagreements among different investigators. Typical findings on age

changes in test performance among children as well as adults will be discussed in subsequent sections.

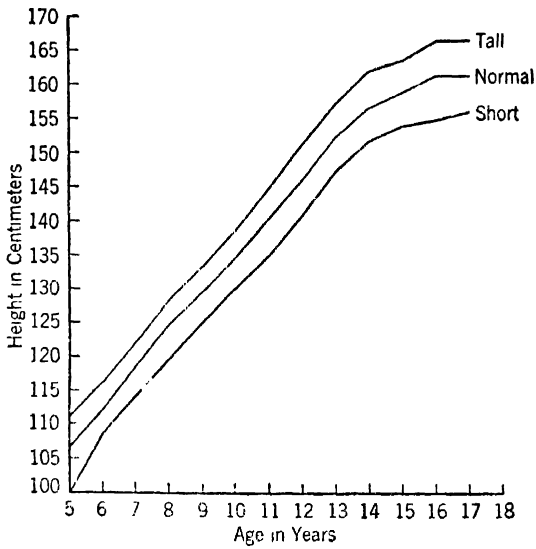


Fig. 48. Growth Curves in Height. (From Baldwin and Stecher, 4, p. 13.)

Cross-Sectional *versus* Longitudinal Comparison. Because the retesting of the same individuals year after year is very time-consuming, growth studies have frequently employed a cross-sectional approach. For example, groups of subjects ranging in age from 12 to 20 are tested simultaneously and the average score of each age group is plotted against age. It is assumed that these averages indicate the normal course of development and that they approximate closely the scores which would have been obtained if, say, the 12-year-olds had been retested annually until they reached age 20.

Such an assumption is open to question for at least some of the groups which have been tested. The different age groups may not be comparable because of progressive *selective factors.* High school seniors, for example, are a more highly select group than high school freshmen, since the poorer students tend to drop out in the course of their high school work. If, as has often been the case, the subjects tested were in school, the higher average score of the older subjects may result in part from this selective dropping out of the less able

students. Had the *same* subjects been tested in the freshman and senior year of high school, the mean gain in score might thus have been much smaller.

A further objection to a cross-sectional approach is that the *experiential backgrounds* of the different age groups may not be comparable. This is especially evident when comparisons are made between widely disparate age groups. For example, the differences between present-day 40-year-olds and present-day 15-year-olds cannot be attributed entirely to factors associated with age. At the time when today's 40-year-olds were 15, schooling was poorer, opportunities for certain types of activity were less frequent or even nonexistent, and many social attitudes were probably quite different from those current today. Such comparisons are thus complicated by the fact that older and younger groups were brought up under different conditions, owing to general cultural changes which are constantly occurring.[1]

Partly in recognition of these difficulties and partly because better facilities for growth studies have become available, an increasing use of longitudinal studies is now being made. To be sure, the longitudinal approach also presents its own peculiar difficulties. Most of these have already been discussed in connection with effects of schooling (cf. Ch. 8). Some of the difficulties, however, do not apply to studies based upon a single group; others are remediable if the investigator is aware of them. Perhaps the most serious weakness of longitudinal growth studies is the somewhat select nature of the participating groups, a fact which results from the prerequisites of stability of residence and continued cooperation with the investigator. At the worst, however, such selection limits the scope of the results, but it does not invalidate them if the population to which they apply is clearly specified.

Among the most extensive longitudinal growth studies may be mentioned the Berkeley Growth Study at the Institute of Child Welfare at Berkeley, California (24); the several Harvard Growth Studies (14, 52); and the ambitious research program of the Samuel S. Fels Research Institute at Yellow Springs, Ohio, which is concerned with nearly every phase of the individual's development from concep-

[1] Cf. e.g., the suggestive findings of Gundlach (19) on the relationship between neuroticism and the socio-economic conditions prevailing at the time when the individual reached maturity. Cf. also Kuhlen (29) for a general discussion of the factor of social change in age comparisons.

tion to maturity (50). Approximately 300 children and their families who live in the neighboring communities are the subjects for the Fels studies. Of the Harvard Growth Studies, the two earlier studies covered only certain aspects of physical growth. The third, completed in 1938, was based upon a wide variety of physical, psychological, and educational measurements of approximately 3500 school children in three Massachusetts cities (14). These children were first tested upon admission to the first grade and were retested annually for 12 years. In the fourth Harvard Growth Study (52), initiated in 1930, testing was begun at the time of birth. Approximately 100 children of each sex are being followed up through periodic examinations in this project.

Average *versus* Individual Curves. A further objection to the cross-sectional approach is that it permits the plotting of only average curves. Since different persons are tested at each age level, it is obviously impossible to chart the progress of individual cases. Even when longitudinal data are collected, moreover, the common practice is to plot the average score for each age. Such a procedure may conceal significant variations from individual to individual. If the development of any particular function varies markedly among different individuals, such differences would probably cancel out in the average curve. The resulting curve might thus be quite unlike the actual course of development for any individual.

A clear-cut illustration of the possible effects of the indiscriminate averaging of individual growth curves is provided by the findings on the *pre-pubertal spurt of growth*. The individual growth curves for many physical traits show a spurt or sudden increase in the rate of growth shortly before puberty. Since individuals differ in the age at which they reach puberty, such a spurt of growth would fall on different portions of the growth curve for different individuals. The curve based upon group averages would therefore reveal no spurt at any period, since this phenomenon would be completely masked or obscured. When only individuals reaching puberty at the same age are included, however, the pre-pubescent growth spurt becomes clearly evident, as illustrated in Figure 49. This curve, taken from the third Harvard Growth Study, shows the average annual increase in height of each of eight groups of girls, classified according to age of puberty. The group reaching puberty before 11½ years, for example, has its maximum spurt of growth at age 11. At the other extreme,

those reaching puberty after 14½ show their most rapid growth at about 14.

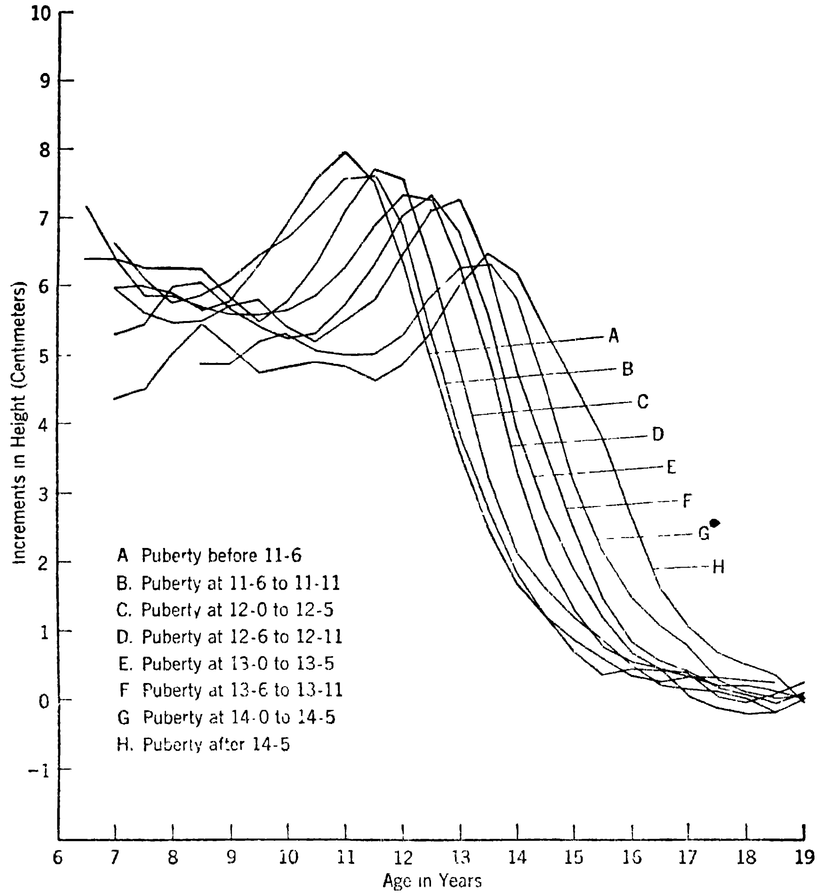


Fig. 49. Average Annual Increments in Standing Height of Eight Groups of Girls Reaching Puberty at Different Ages. (From Shuttleworth, 48, p. 32.)

Difficulty Level of the Test. The form of the growth curve is also affected by several characteristics of the test or measuring instrument employed to gauge the amount of progress. Among such factors is the general difficulty level of the test. In a relatively easy task, performance will improve rapidly during the first few years and

more slowly later on as a perfect score is approached. In a relatively difficult task, on the other hand, or in a task which requires a certain degree of general information or mastery of techniques before it can be properly executed, progress will be slow at first and much more rapid at the upper age levels. The latter task would thus give a positively rather than a negatively accelerated curve. These effects can be illustrated by means of performance curves of successive school grades on easy and difficult sentences in the Trabue Sentence Completion Scale (16). The curves for five representative sentences of different degrees of difficulty are shown in Figure 50.

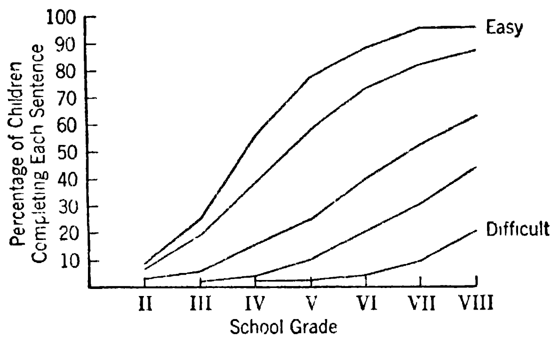


Fig. 50. Grade-Progress Curves for Completion of Sentences of Varying Difficulty. (From Freeman, 16, p. 336.)

Test Ceiling and Test Zero. If the difficulty range of the test is narrow, performance may be artificially cut off at either the upper or lower end, or at both ends. Thus if the *"ceiling"* of the test is too low for the abilities of the older subjects tested, there will not be sufficient items at the difficult end of the scale to permit these subjects to show improvement. Although the subjects' ability may actually increase from, say, 18 to 19 years of age, their scores on such a test may show little or no progress, since their performance is close to a perfect score. The effect which this is likely to have on the form of the growth curve is illustrated schematically in Figure 51. Although the ability which is being measured may continue to increase by equal amounts in successive years (i.e., along a straight line), the scores begin to taper off as they approach the arbitrary test ceiling and will stop rising altogether when the maximum score is reached. An equally artificial slowing down of progress may result at the early

ages from the use of a test whose arbitrary *zero point* is too high for the subjects. Thus if a particular test has too few easy items to sample the performance of the younger subjects adequately, the curve will probably rise slowly at first and then rapidly, i.e., it will be a positively accelerated curve.

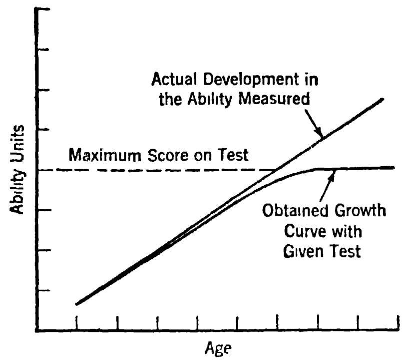


Fig. 51. The Effect of a Low Test Ceiling Upon the Form of the Growth Curve.

Inequality of Test Units. The scores on most psychological tests do not correspond to equal units of ability. The possible effects of such inequalities upon the measurement of improvement with practice were discussed in Chapter 7. The effects upon growth curves are of a similar nature. Let us suppose that on a certain test the difference in ability required to improve from a score of 50 to a score of 51 is considerably greater than that required to progress from 20 to 21. Such a discrepancy would tend to make progress *appear* slower at the later ages, since the 50-to-51 step is more likely to fall within the performance range of the older subjects, and the 20-to-21 step within that of the younger subjects.

A special illustration of the influence of test units upon the form of the growth curve is furnished by *mental age* curves. If average mental age is plotted against chronological age, the result will artificially resemble a straight line. Any divergence from a straight line in such a graph simply indicates errors in test standardization. It will be recalled (cf. Ch. 2) that age scales are so constructed that the average child will advance one year in mental age for each year of chronological age. The successive mental age units are thus adjusted so as to rule out automatically any differences in amount of improvement from year to year. Such units are therefore unsuited to a study of the course of intellectual development.

Type of Abilities Measured at Different Ages. When a complex scale such as the Stanford-Binet is employed, it is likely that different abilities are measured at different age levels. At the upper

ages, most intelligence tests are heavily loaded with verbal functions and other abstract and symbolical tasks. At the other extreme, infant tests are largely based upon sensori-motor development. It is also possible that what appears superficially to be a uniform task may call different activities into play at different age levels. For example, the same form board which measures predominantly spatial perception at age 4 may measure chiefly speed of movement at age 10. It is apparent, therefore, that any one growth curve may in reality consist of *several overlapping curves for different functions.*

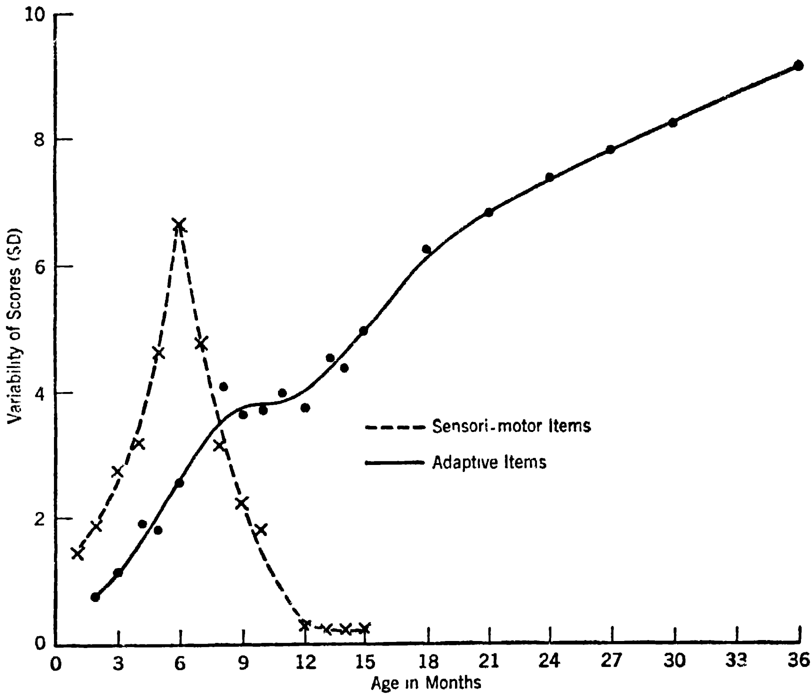


Fig. 52. Age Changes in Variability in Different Functions. (From Bayley, 7, p. 59.)

It is interesting to note in this connection that such shifts from one function to another at different ages seem to be accompanied by rises and falls in *variability,* i.e., in the extent of individual differences. The SD of total scores on infant tests, for example, has been found to rise sharply to a peak at six months of age; it then

drops gradually until the age of one year, after which it rises again slowly (7). It has been suggested that such shifts in the extent of individual differences may parallel the development of separate functions. Thus as the sensori-motor functions approach maturity near the end of the first year, individual differences in these functions — which had previously been large—show a decrease. At this same time, the learning and "adaptive" items, which enter increasingly into the total test performance as the child grows older, introduce a second, gradual rise in variability. The correctness of this explanation of the observed shifts in variability is further indicated by a separate analysis of age changes in SD in the sensori-motor and in the adaptive items. The results of this analysis are presented in Figure 52. It will be noted that the sensori-motor items show a steep rise up to the age of 6 months, followed by a sharp and continued drop, with no subsequent rise. Thus as maturity in these behavior functions is approached, individual differences gradually disappear. The adaptive functions, on the other hand, show a gradual and continuous rise in variability.

Composite Nature of Most Growth Curves. Even at a single age level, the functions involved in most psychological tests are varied and complex. An individual's score on such a test generally depends upon his abilities in a number of different functions. Even if essentially the same functions are measured by such scores over the age range tested, it is nevertheless true that the resulting growth curve is a composite of several curves. Each of the contributing functions may develop at a different rate and reach "maturity" at a different age. To be sure, if the composite is consistently and unambiguously defined, age changes in such a composite may be significant in themselves. The growth curve of height, for example, may be analyzed into separate growth curves for limbs and trunk, which develop at different rates. It is still both practically and scientifically useful, however, to measure age changes in total height. But the composite height measures of different investigators have the same composition—a fact which is clearly not true of different intelligence tests. Many psychological tests purporting to be equivalent may thus yield diverse growth curves, because of the varying combinations of functions which enter into each test.

Three illustrations of such composite behavior indices will be considered. The first deals with the extent of activity in the human fetus at different prenatal ages. Figure 53 shows average results for 16

fetuses observed during normal gestation (49). The top curve, indicating the per cent of time the fetus is active at different ages, resembles the familiar negatively accelerated growth curve. When, however, the total fetal activity is subdivided into the three commonly observed types, three very different curves are obtained. Small, rhythmic movements show little or no increase with age. Kicking shows a sharp rise

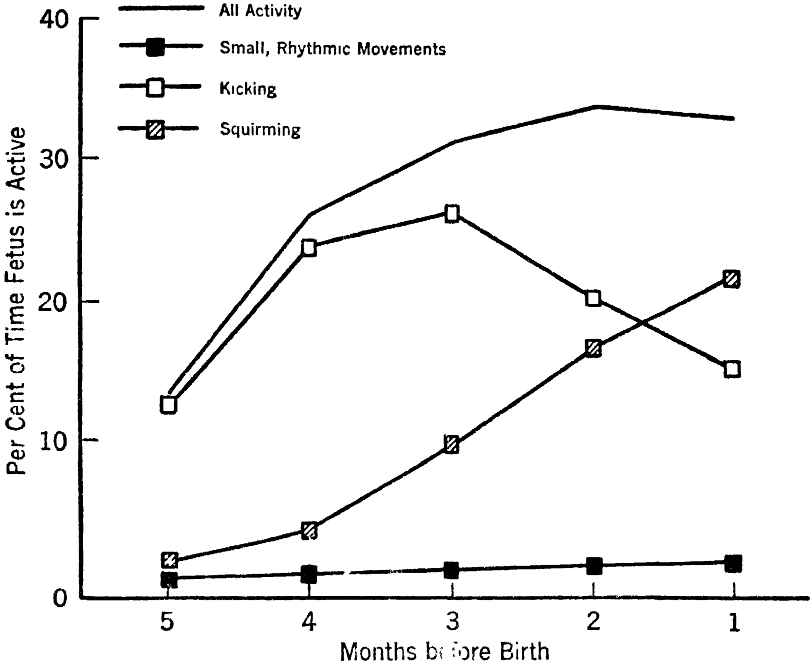


Fig. 53. "Growth Curves" for Different Types of Movement during Prenatal Life. (From Sontag, 49, p. 1, 2.)

from 5 to 3 months prior to birth, followed by a drop. Squirming movements, on the other hand, increase in frequency throughout the observation period, rising slowly at first and more rapidly later.

The second illustration concerns the frequency of crying by infants during the first year of life (6). A record was kept of all instances of crying by 61 infants in the course of monthly physical and mental examinations. The total figures suggest a general tendency for the amount of crying (frequency and duration) to decline until about four months of age, then to increase again, especially after six months.

The amount of crying drops once more beyond six months, but increases slightly toward the end of the first year. On the surface, such a finding might suggest a cyclical development of emotional behavior in the infant. The apparent periodicity, however, may result from the

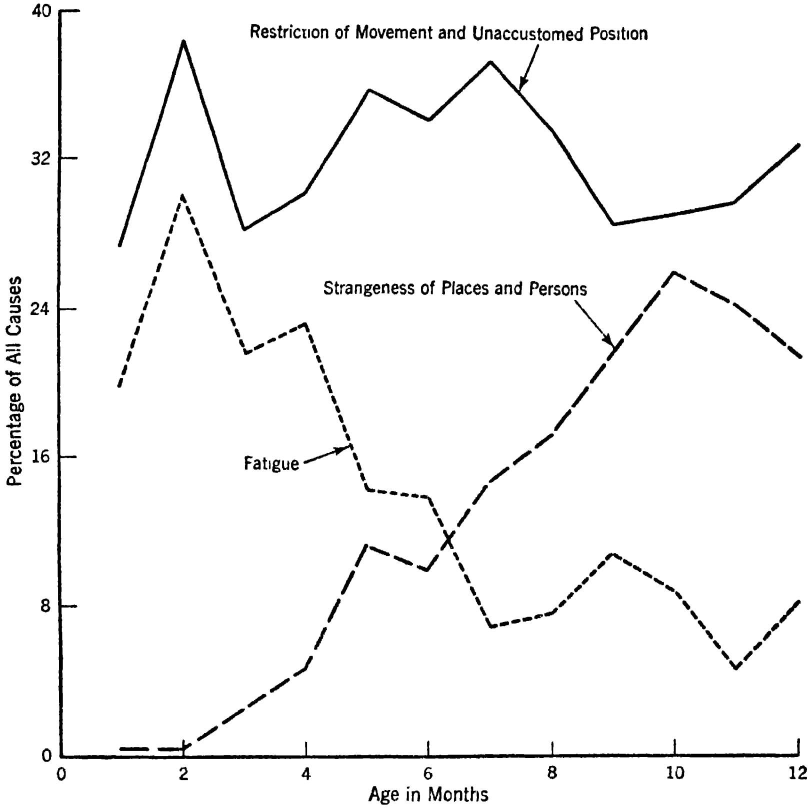


Fig. 54. Age Changes in Crying Behavior in Response to Different Types of Stimuli. (From Bayley, 6, p. 320.)

combined effects of a number of independently varying factors. In the present investigation, crying in response to *different types of stimuli* yielded age curves differing in both form and direction. Three of these curves are reproduced in Figure 54. It will be noted that "crying as a result of restriction of movement and unaccustomed position" retains a relatively high frequency throughout the first year, with no

consistent downward or upward trend. "Crying from fatigue" shows a fairly steady drop from the first to the twelfth month. The reverse trend is evident in "crying because of strangeness of persons or places," which mounts steeply throughout the year. The apparent periodicity in "emotionality" would thus seem to result from a combination of many specific emotional responses, each of which follows its own independent course of development. Such findings suggest that in another investigation a different trend in the composite crying

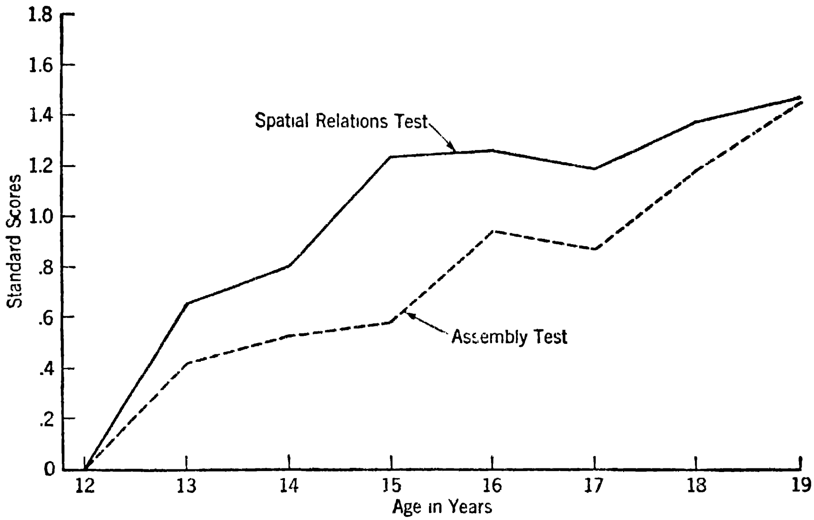


Fig. 55. Age Changes in Two Mechanical Ability Tests. (From Jones and Seashore, 28, p. 141.)

curve might be produced by altering the relative frequency of the specific stimuli which evoke crying.

The third illustration is furnished by the scores made by a group of adolescent boys in the standardization of the Minnesota Mechanical Aptitude Battery (cf. 28). Figure 55 shows the age differences in average standard scores on two of the tests in this battery, the Spatial Relations and the Assembly tests. It will be noted that the curve for Spatial Relations exhibits a definite negative acceleration, rising more sharply until about age 15 and then slowing down. The curve for the Assembly test, on the other hand, follows almost a straight-line trend, with minor fluctuations.

Similar examples could be cited from the various phases of linguistic growth, age changes in different types of memory tests, and the development of many other functions. It should be evident that the so-called curve of mental growth is not one, but many curves. A few of these curves run parallel, others move along simultaneously but at different rates, while still others succeed one another in overlapping steps.

Age Progress Curves. When applied to psychological test scores and other behavior data, the term "growth curve" may be quite misleading. What such a curve actually shows is the performance of the individual at different ages in some standard test situation. Such a curve does not differ in any essential respect from a *learning curve.* In both cases, the subject is tested under similar conditions at successive intervals and his progress is charted on the curve. Learning curves, to be sure, usually cover a shorter period of time than growth curves, although a practice experiment could conceivably extend over several years. The major difference between learning curves and growth curves seems to be that in the former the subject is given special training under rigidly controlled experimental conditions, while in the latter he is left to his own resources. Thus it would seem that a psychological growth curve is at best a practice curve obtained in the absence of controlled conditions.[2] It reflects the cumulative effects of the random training and experience of everyday life, without adding anything essentially new to the picture.

It follows from this discussion that growth curves may vary with the cultural milieu in which they are obtained. If the learning conditions differ from one group to another, the curves of psychological growth may likewise be expected to differ. Such "growth curves" can still serve a useful purpose as descriptive devices. As such they may indicate the general course of development of different functions *under given cultural conditions,* and would characterize individuals of different age levels *within a specific group.* For such curves, the term "age progress curve" would seem a more accurate designation than "growth curve," since it provides a more realistic description of the type of data from which the curves are derived.

[2] This view is to be contrasted with that of Courtis (12, 13), who regards growth as development under *constant* environmental conditions. Although Courtis also considers practice and growth curves as fundamentally similar, he subsumes practice curves under the heading of growth curves, rather than vice versa. It would appear more realistic, however, to regard growth curves as a type of practice curve.

TYPICAL FINDINGS ON THE IMPROVEMENT OF MENTAL TEST PERFORMANCE WITH AGE

In view of the many difficulties enumerated in the preceding section, the reader may wonder what, if anything, can be learned from an examination of the age curves themselves. First, it is clear that such curves show changes only in the particular area of behavior which the

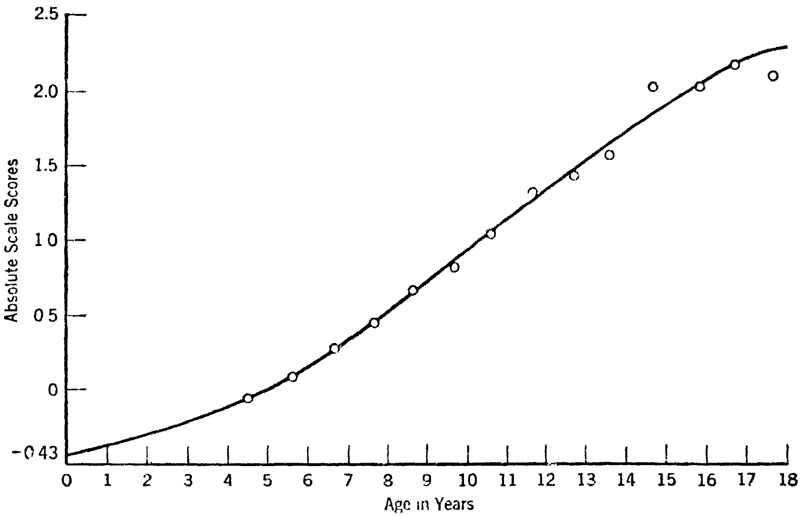


Fig. 56. Age Changes in Stanford-Binet Performance Expressed in Absolute Scale Units. (From Thurstone and Ackerson, 62, p. 576.)

test samples. Secondly, the results cannot be assumed to hold for individuals whose experiential background is markedly unlike that of the group on whom the curve was derived. Thirdly, the units in which test scores are expressed and plotted present a persistent problem, to which various alternative solutions have been proposed. Certainly, some type of equal-unit score would seem to permit a more intelligible picture of annual progress than is furnished by unequal raw scores or by artificially adjusted mental age units. With these points in mind, we may examine the results of some of the most carefully conducted investigations.

Perhaps the most widely quoted curve is that prepared by Thurstone and Ackerson (62) from the Stanford-Binet scores of 4208 sub-

jects between the ages of 3 and 18. The scores were first converted into an equal-unit scale.[3] The resulting curve is reproduced in Figure 56. It will be noted that this curve rises slowly at first, then more rapidly, and then slowly again as the final leveling-off is approached. Thus the curve is described as positively accelerated in its early stages and negatively accelerated later on. The portion of the curve extend-

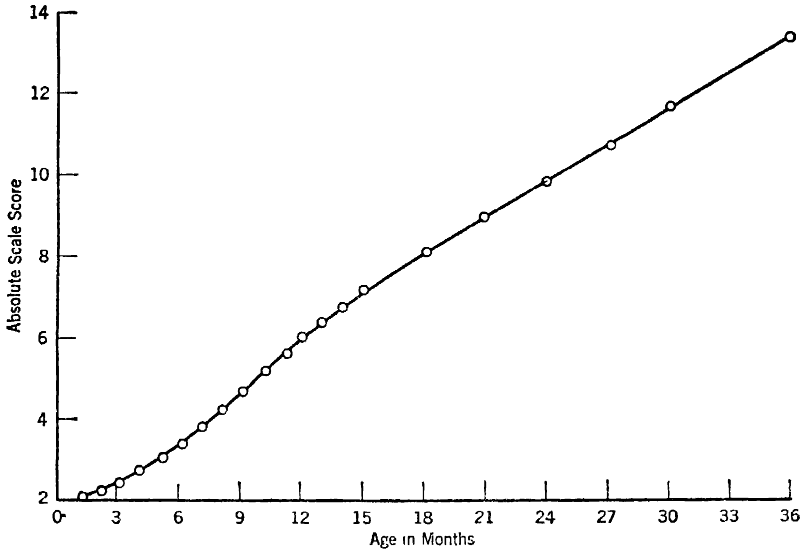


Fig. 57. Age Changes in Test Performance during Infancy, Expressed in Absolute Scale Scores. (From Bayley, 7, p. 43.)

ing below three years of age was found by extrapolation. Some confirmation of its shape, however, was provided by the data of Bayley (7), which were plotted by Thurstone in the same equal-unit scale. This curve, based upon retests of infants in the Berkeley Growth Study from shortly after birth to age 3, is shown in Figure 57. It will be seen that the curve shows positive acceleration, especially during the first seven or eight months of life.

A curve similar in general form to that of Thurstone and Ackerson was found in the third Harvard Growth Study, through retests of 522 children between the ages of 8 and 15 (cf. Fig. 58). The findings of the earlier cross-sectional study are thus confirmed by a longitudinal approach. In another longitudinal study, Freeman and Flory (17)

[3] "Absolute scale" units, whose derivation can be found in Thurstone and Ackerson (62).

retested 469 children annually with a battery of four tests. The testing began at age 8 and extended over a ten-year period, more than half of the subjects receiving five or more successive retests. The test battery (designated VACO) consisted of vocabulary, analogies, completion, and opposites. The composite curve showed an almost linear, or uniform, increase from ages 8 to 15, with only a slight decline in

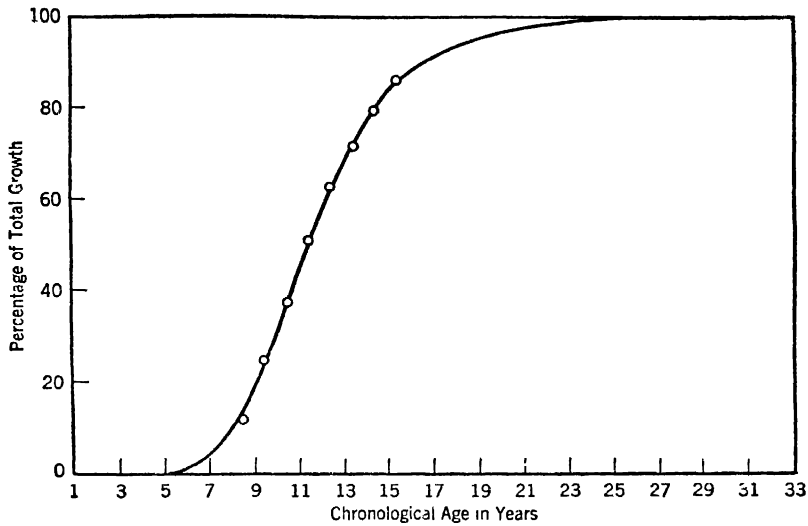


Fig. 58. Age Changes in Intelligence Test Scores in a Group of 522 Children Retested between the Ages of Eight and Fifteen. (From Dearborn and Rothney, 14, p. 215.)

rate beyond that age. Considerable progress was noted even at the highest ages tested. Pintner and Stanton (44) administered the CAVD examination annually to 140 children in grades I to VIII, each child being tested over a period of from two to six years. This test consists of four parts—completion, arithmetic, vocabulary, and directions—which recur at each of the difficulty levels. A particular advantage of the CAVD for such a study is that its units progress by steps of equal difficulty. The average progress curve obtained in this study showed some negative acceleration, thus confirming the usual results for these age ranges. The mean annual gain dropped from 1.23 between the ages of 7 and 8 to .60 between the ages of 13 and 14.

In conclusion, it may be noted that when short periods of five years or so are studied, especially during middle childhood, the age progress

curve for most "intelligence" tests appears to be approximately a straight line. Negative acceleration, or slowing down, sets in as the leveling-off point is approached. In other words, cessation of progress occurs gradually rather than suddenly. As for the lower end of the curve, in infancy and early childhood, its form is still rather uncertain, owing to insufficient empirical data. Finally, it should be added that many investigators have found marked *individual differences* in the form of the "growth curve." Cycles of slow or rapid development, which may cover several years, occur in many individual curves, with no regularity from person to person.[4] Evidently the detailed course of behavior development is influenced by many factors—structural and experiential—which vary with the individual.

ADULT INTELLIGENCE

The study of maturity and old age is a very recent but rapidly growing branch of psychology. Such interest in the characteristics and the problems of older persons has taken many forms. A number of research projects on fairly large groups of adults have been concerned with changes in intelligence, special aptitudes, or emotional characteristics. Special efforts have been made to study representative samples at the various age levels, in contrast to the rather atypical groups in the earlier studies on older persons. An increasing concern with the vocational and personal guidance of older persons and with the clinical treatment of maladjustments at these age levels is also noticeable. Several books, round-table discussions, and even a special division of the American Psychological Association devoted to problems of maturity and old age furnish further evidence of the status of this area in contemporary psychology (cf. 27, 30, 31, 32, 42). Mention may also be made of the construction of intelligence tests specially designed for adult groups (20, 63).

Limit of "Mental Growth." The more recent and better controlled studies have generally found that intelligence test performance tends to improve until the very late teens or early twenties. Some evidence on this question is furnished by test norms, although when the tests are standardized on school populations, selective elimina-

[4] Cf., e.g., Bayley (8), Dearborn and Rothney (14), Freeman and Flory (17), and Jones and Conrad (26).

tion is likely to make the norms spuriously high at the upper ages. In the Terman-McNemar Test of Mental Ability (55), the norms were corrected for such selection and can therefore be regarded as more nearly representative of the performance of comparable groups at each age. The average annual gain in standard score on this test dropped gradually from 7 points between the ages of 10 and 11 to 3 points between the ages of 18 and 19. It is evident that, although the annual increments tend to diminish, performance is still improving at age 19.

Other cross-sectional comparisons were made by Teagarden (54) on an orphanage sampling and by Jones and Conrad (25) on the entire population of certain New England villages. The results of these two studies also show gains in test scores up to age 18 and probably even later. In a survey by Miles and Miles (41) on groups ranging in age from 7 to 94, the Otis Intelligence Test scores continued to rise until about age 20. This was also the approximate age at which improvement ceased in the standardization data of the Wechsler-Bellevue Intelligence Scale (63).

Longitudinal studies on the same individuals have likewise furnished evidence that intelligence test scores continue to improve until the age of about 20 (cf. 17, 56). In one of these studies (17), some of the subjects who had been first tested at nine years of age were followed into college and were still making significant gains at the termination of the survey. In this connection may also be cited the retest studies on high school and college students with parallel forms of the Psychological Examination of the American Council on Education (5, 23, 37, 61). Large and consistent gains upon retesting were found in all these studies, not only in average score, but also for nearly every individual. Without the use of non-college control groups, it is of course impossible to determine the extent to which such gains are attributable to college training and to other more general conditions. From a purely descriptive viewpoint, however, the fact remains that gains in intelligence test performance were made consistently by these subjects, who ranged in age from the middle teens to the late twenties. There is some evidence, moreover, that individuals who continue their education longer tend to improve in intelligence test performance until a later age.

Course and Amount of Decline. Closely related to the question of "intellectual maturity" is that of the subsequent decline of abilities

with increasing age. A few investigations have been specifically directed to this problem. In one of the first systematic surveys of the intelligence of older persons, Jones and Conrad (25) gave the Army Alpha to 1191 persons between the ages of 10 and 60, constituting nearly the entire population of 19 villages in rural New England. Miles and Miles (41), also in the effort to obtain roughly comparable samplings at different age levels, obtained adult subjects through lodges and social groups. Their total sample consisted of 823 subjects ranging in age from 7 to 94, all of whom were given a shortened form of the Otis Self-Administering Intelligence Test. In the process of standardizing the Wechsler-Bellevue Intelligence Scale, Wechsler (63) obtained scores from 670 children and 1081 adults ranging up to 69 years of age. All the subjects in this survey lived in New York State, the adults being selected so that the occupational distribution for each age level resembled roughly the corresponding distribution in the national census data.

The age curves obtained in these three investigations are given in Figure 59, each curve showing the general trend of average scores with age. In order to make the data of the three investigations comparable, scores on all three tests were first converted into standard scores (cf. 27). It will be noted that rate of decline is steeper in the Wechsler and in the Miles and Miles studies than in that of Jones and Conrad. This difference may be in part the result of sampling irregularities at the upper ages. Differences among the tests employed in the three studies, however, are probably a major factor.

The consideration of these average trends must be qualified by the fact that *individual differences* were large at all ages, with extensive overlapping of different age groups. Variability, in fact, tended to increase with age, despite the decline in average scores. Such increases in variability are especially noteworthy in view of the possible effect of selection in the older age groups. Thus among the oldest groups, more individuals pleaded exemption from the tests because of failing eyesight, reading difficulties, and the like. Moreover, the less energetic and possibly less intelligent older persons were less likely to come to the centers where the tests were given. In the Jones-Conrad study, for example, a sharper age decline was found among the subjects tested in their own homes than among those tested in community centers. These selective factors would tend to reduce

variability at the upper ages, by elimination of individuals at the lower end of the distribution. Hence if all subjects had been tested, the obtained rise in variability would probably have been still greater. It would seem that, as the individual's experiential background is enriched with age, more sources of variation in behavior are introduced and individual differences continue to increase.

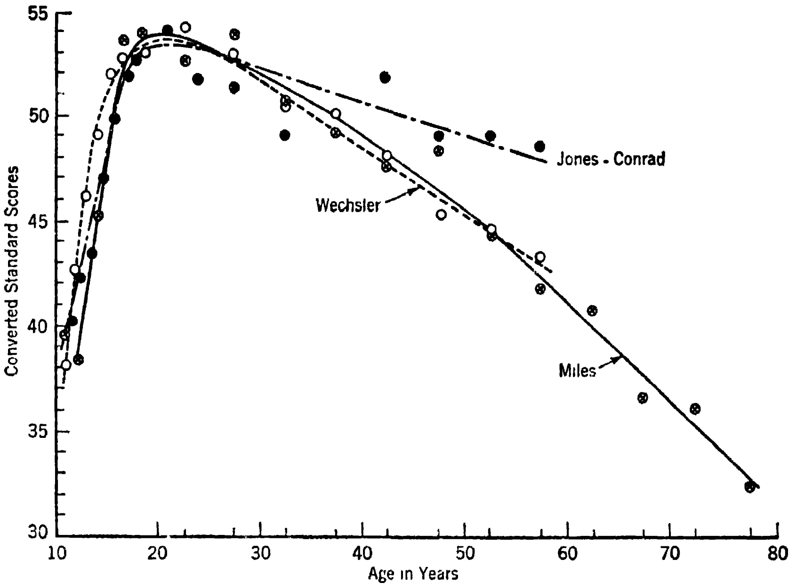


Fig. 59. Intelligence Test Score in Relation to Age: A Comparison of Results from Three Investigations. (From Jones and Kaplan, 27, p. 72.)

The *overlapping* of the various age groups is such that individual differences within any one age level are much larger than the differences between age groups. Thus the brightest persons in even the oldest groups tested were still conspicuously better than the dullest persons in the younger groups. Further corroboration of such a finding is provided by a study on a small sampling of persons of uniformly high educational level (53). In a comparison of the psychological test performance of 45 university professors, aged 60 to 80, with a comparable group of 45 academic men between the ages of 25 and

35, individual differences were again found to be much more impressive than age differences.

That age in itself is a poor guide to ability level is further illustrated in Figure 60, which is based upon an analysis of the Miles and Miles data. The adult subjects were classified into four levels

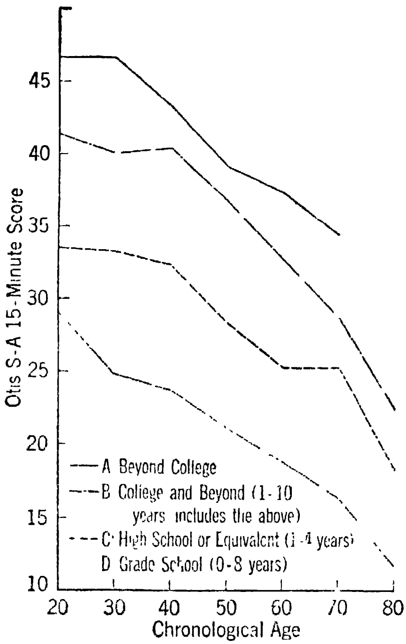


Fig. 60. Age Changes in Intelligence Test Scores at Different Educational Levels. (From Miles and Miles, 41, p 70.)

in terms of the amount of formal education they had received. The highest level (A) consisted of college graduates who had received additional professional or graduate training; the lowest level (D) extended from a total lack of formal education to elementary school graduation. Although all four groups show a decline in mean Otis score with age, the four curves neither cross nor meet. In other words, the higher educational groups retain their superiority consistently at all ages. It should also be noted that the lowest point on curve A, reached by the 70-year group, is still higher than the highest points of curves C or D. Thus a 70-year-old person who had pursued at least

one year of graduate work would be expected to score higher than a 20-year-old elementary school graduate.

One of the most significant findings of adult testing has been the *specificity of age changes.* The curve of decline, even more than the curve of growth, varies with the type of ability measured. Some abilities increase with age, some decline, others show little trend in either direction (cf., e.g., 51). An illustration of this fact is to be found

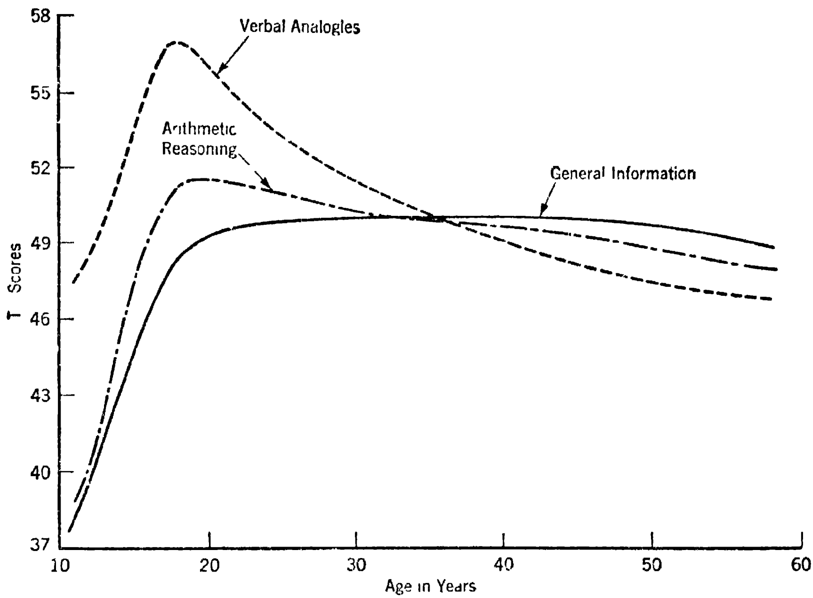


Fig. 61. Decline of Ability in Different Functions. (From Jones and Conrad, 25, p. 250.)

in the Jones-Conrad study. When scores on each of the sub-tests of the Army Alpha were plotted against age, several dissimilar curves were obtained. Three of these curves, for the tests of general information, arithmetic reasoning, and verbal analogies, respectively, are reproduced in Figure 61. The continued rise in general information beyond age 20 is typical of findings by other investigators, as is the slight decline in arithmetic reasoning. The steeper drop in verbal analogies may result in part from the relative prominence of speed in this particular sub-test and in part from the nature of the test

itself.[5] In another investigation (21), a vocabulary and three performance tests [6] were administered to 375 men and 268 women of low socio-economic level, ranging in age from 15 to 76. The vocabulary test showed a rapid rise with age from 15 to 20, and slower rises up to 55, when a slight drop was found. The scores on the three performance tests, on the other hand, declined rapidly at the older ages. The greater stability of vocabulary tests among older persons, as contrasted to the rapid decline on many other tests, has been corroborated by a number of investigators (cf. 3, 47, 64).

Tests which emphasize *speed* have regularly shown a particularly steep decline with age. This finding has been reported consistently in investigations in which the performance of younger and older adults was compared on power as well as on speed tests (cf. 32, 39, 42). For example, when subjects ranging in age from 15 to 75 were given the Otis Self-Administering Intelligence Examination by a work-limit rather than a time-limit method, the younger adults required less time to complete the test, but the accuracy scores showed virtually no relation to age (11). In another study (38), three groups of subjects, aged 20 to 25, 27½ to 37½, and 40 to over 70, respectively, were matched in CAVD scores. It will be recalled that this is a pure power test, given with practically no time limit. When the three equated groups (totaling 143 cases) were tested with the Army Alpha and the Otis, the average scores on these tests dropped progressively from the youngest to the oldest group. Speed plays an important part in performance on the latter two tests.

A further point to note in evaluating the age decrements in mental test scores is that nearly all our available tests have been *constructed on young people*. As a result, they are overloaded with tasks typical of the activities of young persons (cf. 20, 42). The content might be quite different if the tests had been constructed by sampling the activities of older persons. It is likely, therefore, that most available tests tend to favor young persons unduly. The fact, for example, that most intelligence tests are so heavily weighted with academic content reflects the school activities of the young subjects on whom they were

[5] It will be noted in Figure 61 that the general level of scores in the three tests differs, despite the use of T-scores in plotting all these curves. This results from the fact that the T-scaling was done on the basis of the scores of a sub-sample of 200 adults aged 25–39, rather than on the basis of the entire sampling.

[6] Knox Cubes, Porteus Mazes, and Ferguson Form Boards.

standardized. The inferior performance of older groups on such tests may result partly from the greater distance in time of the older person from his own formal education. The loss would thus be more a matter of forgetting and lack of practice than one of deterioriation. Moreover, in a number of comparisons in which level of education was not held constant, it is probable that the older persons had received less education than the younger, since the general level of education has risen considerably during the past two or three generations.

Adult Learning. The notion that "you cannot teach an old dog new tricks" is a common one in popular thinking. Adults frequently deplore their inability to learn a new language, a new motor skill, or an improved work method as well as they could in their younger days. Closer observation reveals, however, that the *conditions of learning* are far from comparable at different age levels. The time available for learning, the distractions, and the motivation for learning are often very different for the child and the adult. The learning of new skills is frequently undertaken casually and halfheartedly by the adult, while for the child or adolescent it is the core of his serious responsibilities, other activities being "extracurricular."

When older and younger persons learn under comparable conditions in an experimental situation, the differences in their performance are relatively slight. In a series of investigations with a variety of tasks (57), an average decline of less than 1% a year in "sheer modifiability" was found between the ages of 22 and 42. This decline was manifested principally in the more meaningless tests of rote learning, such as drawing lines of given lengths blindfolded, learning a code, or memorizing numbers paired with nonsense syllables. In most other tasks, the older persons could compensate for any loss in learning ability by greater interest, better sustained effort, and a larger fund of relevant experience. For example, in stenography and typewriting, in learning Esperanto, or in university courses, the progress of the older persons equalled and sometimes even excelled that of the younger.

When the new learning runs counter to previous learning, it is reasonable to expect older adults to be handicapped. This is a simple result of *interference*, or "negative transfer," which has no necessary connection with age as such. There is some experimental evidence

(46) to show that tasks which are hindered by previous experience suffer a greater age decrement than those which are benefited by such experience. In a learning experiment involving the comparison of three different age groups, the older subjects were found to be inferior to the younger subjects in the learning of all types of material, although they were *less* inferior on the more meaningful material. For example, the older subjects did relatively best in learning paired associates which were meaningfully connected, such as nest-owl, soft-chair. Their performance was poorer in learning "nonsense" material, such as $A \times M = B$ or $N \times M = C$, and poorest in memorizing material which conflicted with previous learning, such as $3 \times 4 = 2$ or $3 \times 1 = 1$. Some of the attitudes and emotional reactions characteristic of "old age" may have a similar explanation (30). The greater "conservatism" commonly attributed to older persons may simply mean that the longer one has held a certain opinion, in general, the more firmly fixed it becomes. Older persons, for instance, were found to be less susceptible than younger persons to suggestion from either group opinion or the opinion of experts (40). Such reactions may be explicable in terms of the cumulative effects of previous experience, without resort to unknown physiological bases. Through all this consideration of adult learning, we must not, moreover, lose sight of the wide individual differences and extensive overlapping of age groups, as conspicuous in this area as in any other age comparison.

The Age of "Maximum Productivity." Another approach to the study of adult abilities has been through an analysis of productive or creative work in such fields as science, literature, and art (33, 34, 35). To be sure, selected cases can be found to illustrate maximum productivity at almost any adult age in individual scientists, inventors, writers, musicians, or artists. In terms of group trends, however, fairly consistent age curves have been found. In Figure 62, for example, is shown the average number of "best books" by 101 noted writers during successive five-year periods of their lives. Consistently similar curves were obtained when the lists of "best books" prepared by different accepted authorities were consulted. It will be noted that the most "productive" years for such writers fall between the ages of 40 and 50. Similar surveys among scientists and inventors, on the other hand, showed maximum productivity to occur between 25 and 40, with a subsequent dropping off in later years. The specificity of such

trends is further illustrated by the finding that the peak of production among musicians differs with the type of music. Similarly, among writers the peak occurs earlier for poets and later for writers of historical, critical, philosophical, or scientific works.

Although of interest in themselves, such data on productivity do not tell us very much about the rise and fall of abilities in general. In the first place, the subjects are certainly a highly selected group and not typical of the general population. The possibility that the age of maximum production does not coincide with the age of qualitatively best production for each individual must also be taken into account. It may well be that at certain ages quantity is sometimes sacrificed for

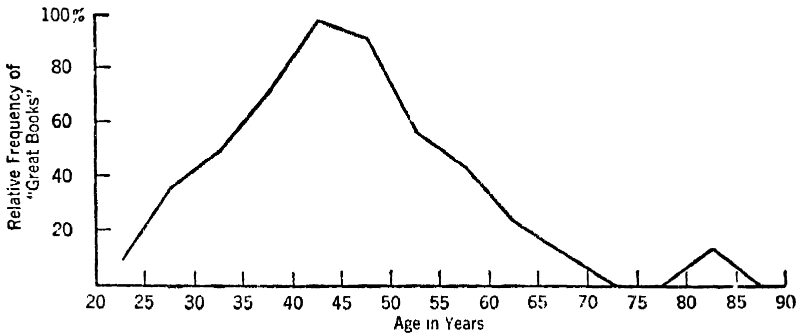


Fig. 62. Age Changes in Literary Production. (From Lehman, 34, p. 66.)

the sake of quality in creative work. Thus in a survey of the age at which over 4000 scientists produced their "chief work," the median age was found to be 43 years (1). This is considerably older than the "age of maximum productivity" reported above for scientists.

A possible decrease in motivation because of financial and professional security, development of other interests, and the like, may also affect productivity among some older persons. Closely related to this factor is the commonly noted increase in administrative duties with increasing age, especially among academic persons (cf., e.g., 9). Such responsibilities sometimes seriously interfere with "creative activities." Finally, the results of such surveys may be specific to the particular historical period covered and may vary as social conditions

vary. A recent analysis has shown, for example, that present-day leaders in a number of fields are definitely older than were their predecessors who held the same nominal positions (36).

THE CONSTANCY OF THE IQ

The widely debated question of the constancy of the IQ at different ages can be better clarified if it is regarded as two separate questions. The first is the purely empirical, practical, "actuarial" question of *prediction*. It is generally recognized that the intellectually gifted school child is likely to develop into a superior adult, and that a feebleminded child will probably fall below average as an adult. Just how accurately can such predictions be made, and how early in the individual's life? These are the practical questions of prediction, concerned only with observed trends and regularities. The second question is a theoretical one, in which the degree of constancy of the IQ is considered as an index of the *regularity of mental development*. We shall see that the answer to the first question does not necessarily imply a corresponding answer to the second.

Empirically, the IQ has been found to remain sufficiently constant during the elementary school years to make prediction over several years feasible. Among older subjects, intellectual level likewise shows considerable stability, especially when individuals remain in fairly constant environments. Thus the intelligence test scores of college students correlate very highly with the scores obtained by the same individuals in high school or even in the upper elementary school grades (15, 60). In one study, for example, a correlation of .80 was found between scores on the American Council Psychological Examination administered at college entrance and intelligence test scores obtained as early as the seventh grade of elementary school (15, p. 476). It is of the utmost importance, in interpreting such results, to realize that only subjects who had continued their education to the college level were included. If the investigators had worked from the other end, by following up a group of elementary school graduates and retesting them after five or six years, the correlations would probably have been much lower, since the intervening educational and other experiences of the subjects would have undoubtedly varied much more widely. In a group with comparatively constant educational

experiences, however, individuals tend to maintain the same relative position in intelligence test score over a period of many years.[7]

In general, there are two major exceptions to the constancy of the IQ. First, large shifts in IQ may occur among individuals who have undergone fairly drastic environmental changes, such as placement in a foster home or participation in a specially designed and intensive remedial program (cf. Ch. 8). Secondly, preschool tests have proved to be of little or no value in predicting IQ's in adolescence and adulthood.

Evidence for the latter finding is plentiful. For example, in a group of 123 children, performance on the Gesell schedule at 6 months correlated only .37 with Merrill-Palmer scores at age 2 (43). In the same group, a correlation of .46 was found between the initial Gesell test and the Stanford-Binet IQ at age 3. In the course of the Berkeley Growth Study, 61 children were retested regularly from the age of one month to 9 years. From an analysis of their scores, Bayley (8) concluded that available intelligence tests for infants and young children cannot be used to predict later ability. Tests given at age 4 may permit grade school predictions within wide classifications; tests between 2 and 4 will predict 8- or 9-year performance with some success; but scores obtained before 18 months of age are completely useless in the prediction of abilities during school ages (8).

In another study of 252 children participating in the Berkeley survey, Honzik (22) likewise found little prediction possible from early tests. Initial tests made at 21 months correlated only about .30 with retests at 5 and 6 years of age. Somewhat higher predictive value was shown by tests in the upper preschool ages, but the correlations were still too low for individual estimates. Essentially the same conclusion was reached by Goodenough and Maurer (18) in follow-ups of over 200 children who had taken the Minnesota Preschool Test before the age of 6. Correlations of these initial scores with Stanford-Binet retests at ages 7 to 12 ranged from .15 to .45. Correlations on smaller groups who were followed into college were also reported. The correlations between preschool tests and scores on the A.C.E. examination

[7] The fact that college students represent a highly select group intellectually, and are therefore more homogeneous in intelligence test scores than a random sampling of the general population, would tend to *lower* the correlation between initial and terminal scores in a college group. The high correlations actually obtained thus indicate even more vividly the effect of the continued uniformities in these subjects' educational experiences.

taken upon college entrance were .12 (with tests taken under age 4), .29 (with tests taken between ages 4 and 5), and .39 (with tests between ages 5 and 6).

Several investigators agree that the low predictive value of infant and preschool tests cannot be attributed entirely to the unreliability of such tests, since reliability coefficients found within short periods are often quite high. Moreover, high and low scores tend to occur in clusters within any one individual's successive retests, and thus seem to indicate periods of lag or spurt which may extend over several years. At least two other explanations have been suggested. First, the individual's development may be more susceptible to environmental influences at early ages. Secondly, different types or combinations of functions may be covered by preschool and by subsequent "intelligence" tests. Some evidence for the latter explanation has already been cited in earlier sections. It is probable that both factors contribute to the low predictive value of early tests.[8]

Retest correlations also depend upon the *interval between retests.* In other words, the interval over which predictions are made affects the accuracy of the prediction. This relationship is clearly demonstrated in an analysis conducted by R. L. Thorndike (58) with previously published data on school-age children. By combining the results from those studies with fairly uniform test-retest intervals and then fitting a curve to these data, Thorndike obtained an equation showing the relationship between time interval and expected correlation. On this basis he estimated, for example, that the test-retest correlation is .90 for an immediate retest, but drops to .70 over a five-year interval. The correlations empirically obtained with school children by subsequent investigators have in general corroborated the values predicted from this curve (cf., 10, 59).

To recapitulate, the predictive value or consistency of intelligence test scores increases with the *age* at which the test is administered, and decreases as the *interval* between test and retest increases.[9] Both of these relationships can be explained on the basis of the "*overlap*" of

[8] An additional explanation, in terms of "overlap," will be discussed below.

[9] It should be noted that the effect of length of test-retest interval upon constancy of the IQ will itself vary with age. At the older ages, the same time interval is accompanied by a much smaller change in test performance, and even relatively long intervals are likely to yield fairly stable results. Thus in the case of tests administered to college students, correlations with earlier tests taken during the freshman year of high school are about as high as those with tests taken during the senior year of high school (60).

abilities at successive age levels (cf. 2). The performance of the older individual is based in part upon his retention of abilities which he manifested at earlier ages. The older the individual, the greater the proportion of such overlap between present and earlier performance. Let us consider a simplified, schematic illustration. A child's IQ at age 3 may be determined by his successful completion of 10 items which he was also able to pass at age 2, plus 5 additional items; in this case the overlap of his performance at ages 2 and 3 is 10/15 or 67%. On the other hand, at age 16, this child's IQ may depend upon 48 items which he was also able to pass at age 15, plus two new ones; now the overlap is 48/50 or 96%.

In his presentation of this "overlap" hypothesis, J. E. Anderson (2) has summarized the relationship as follows:

> We deal here with a phenomenon in which the prediction of final status is based upon a larger and larger proportion of that which is included in the total; that is, scores at 10 years include more of that which is present at 16 years than do scores at 3 years. . . . Since the growing individual does not lose what he already has, the constancy of the IQ is in large measure a matter of the part-whole or overlap relation (2, pp. 388–394).

In support of this hypothesis, Anderson computed a series of correlations between initial and terminal "scores" obtained with shuffled cards and random numbers. These correlations, which depended solely upon the extent of overlap between successive measures, agreed closely with test-retest correlations in intelligence test scores found in three published longitudinal studies. In fact, the test scores tended to give somewhat *lower* correlations, a difference attributed by Anderson to such factors as errors of measurement and change in test content with age.

Further corroboration of this explanation of the constancy of the IQ in terms of overlap is furnished by an analysis by Roff (45). Using previously published data, Roff correlated the intelligence test performance of children at any one specific age with their gain in performance after one or more years. These correlations were all close to zero. From such a finding, the author concludes that "the so-called 'constancy of the IQ' is due primarily to the retention by each child of the skills and knowledge which determined his scores in earlier years, and is not due at all to correlation between earlier scores and later gains or increments" (45, p. 385). These findings provide

an answer to the second question regarding the constancy of the IQ, viz., does the empirically observed constancy of intellectual status signify *regularity of mental development*? The answer now appears to be clearly "No." The growing individual exhibits an increasing consistency of ability level, not because the "rate of growth" is constant, but because his present accomplishments constitute an ever increasing portion of his future accomplishment as he grows older. This is tantamount to saying that at age 15 we can make a more accurate prediction of an individual's subsequent behavior than at age 2, because we know more about him at 15. The proportional change in his behavior from age 15 to 16 is less than from age 2 to 3, and certainly much less than from 2 to 16.

TRAINING AND GROWTH

We may now attempt to synthesize the findings of the various investigations and to evaluate them in the light of the studies on training discussed in earlier chapters. If we think of mental development in terms of *learning*, the diverse findings both on the upper limit of mental growth and on the decline of ability can be fitted into an intelligible pattern. It might be objected that the learning curve shows no decline, whereas age curves do. This apparent inconsistency results, however, from an incomplete statement of the situation. The problem will be considerably clarified if we speak of age changes in specific tasks, as we do in the case of learning, rather than discussing mental development in general. It is quite true that the cumulative effects of learning in everyday life will increase proficiency indefinitely in certain tasks, but such learning will just as surely interfere with the performance of other tasks. If the general effect of any specific act of learning upon all the individual's behavior is considered, it becomes apparent that learning may cause a decline as well as a rise in achievement.

The decline in performance on most psychological tests with age is no longer surprising when we realize the resemblance of all such tests to school work. We should therefore expect that the longer the individual has been out of school, the more chance he has had to forget what he learned as a child, *through interference from other activities*.

Although in his everyday life the adult may be employing much that he learned in school, he is at the same time losing many school

habits, such as working with a specific time limit, following directions literally although he may see little sense in them, and working with materials which may be meaningless and of no apparent use to him. When a school child is confronted with a psychological test, the novelty, strangeness, and apparent purposelessness of many of the things he is asked to do will not disturb him unduly, since at that age he is still doing many things for which he can see no immediate value. Such tasks are accepted by the child as part of his everyday work. Not so with the adult. The older he grows, the more he concentrates only on those activities which are either of practical significance or directly pleasurable to him. The reaction of many adults to intelligence tests, as contrasted with that of school children, illustrates this difference. To most adults, such a test is either foolish or entertaining. The adult is far more sensitive to the apparent impracticality of the situation than the child, who is accustomed to taking tests which to him may seem equally useless.

That adult ability does not decline in all tasks is demonstrated by the obvious improvement in functions related to the individual's daily work. The achievements of many people progress along a continuously rising line throughout life.

Nor can a distinction be legitimately made between the *extent* of a person's abilities, which increases constantly with age, and the *level* or difficulty of task which he is capable of mastering. The latter is definitely dependent upon the former. As was brought out in the discussion of practice and variability, the more an individual has learned, the better able he is to lear. A problem which is commonly regarded as difficult and which can be solved by only a few individuals is often one which involves the synthesis of more numerous and varied types of learned behavior. We can say, for example, that the derivation of a formula which requires a knowledge of arithmetic, algebra, trigonometry, and calculus is more difficult than one which can be derived simply by the application of principles of arithmetic and algebra. If we define the difficulty of a task objectively in terms of the number of people who can perform it correctly it will unquestionably prove to be related to the number of different specific abilities involved. Even if a more subjective, popular definition of difficulty were suggested, it would doubtlessly be found to hinge upon the same principle.

Many of the previously reported findings are clarified if we consider mental development in terms of learning. Thus the limit of intel-

lectual improvement, as measured by common intelligence tests, will be reached later by those groups which continue their formal schooling to a later age. This has been repeatedly demonstrated in studies on the "point of cessation" of intellectual growth. The data on the constancy of the IQ, apart from the purely statistical influence of "overlap," are also in general conformity with such a "learning hypothesis." It is during the elementary school years that *predictions* of subsequent performance can be most accurately made for individual subjects. These are just the years when American school children, upon whom these studies were conducted, undergo the standardized intellectual experiences provided by uniform curricula. During the preschool years and again in adulthood, individuals' experiences are less standardized, and their intellectual performance is less predictable.

Corroboration of the proposed interpretation of age changes in mental traits can also be found in the experiments on adult learning. It will be recalled that the rate of decline was more rapid for the more "meaningless" than for the more "meaningful" and useful tasks. A similar difference was found between those tasks which were hindered and those which were aided by the common training furnished in our culture.

Is there any physiologically determined decline in mental activity with age, apart from the changes related to learning? The effect of the deterioration of requisite structures doubtlessly plays a part in the marked and sharp decline in all psychological functions which frequently characterizes senescence. Such obvious handicaps as failing vision and hearing, and muscular and neural deterioration can hardly fail to affect all the individual's activities. These changes, however, do not set in to an appreciable extent until very late in life, and consequently cannot plausibly be offered as an explanation of the decline in mental test performance during earlier maturity. There are persons, moreover, in whom serious structural handicaps during old age have been compensated to a remarkable degree by interest, effort, and the advantages of past experience. The wide individual differences found within any one age level also suggest the importance of specific environmental circumstances.

The physical handicaps of senescence may be regarded in the same light as the physical inadequacies of the immature child. Both set the upper limits of behavior development at a given chronological period, but they do not determine the degree to which such limits will be

approximated. It seems, also, that these physically set limits are always much higher than is commonly suspected, since training and stimulating conditions can at all ages accomplish surprising results. Finally, it may be added that age changes in behavior may also vary in different cultures (or cultural sub-groups) in which the attitude toward old age differs.

CHAPTER

10

Family Resemblance

THE INTERPRETATION OF FAMILY RESEMBLANCES is complicated by the fact that close relatives generally live together. The environment of individuals within a single home is certainly more similar than in any other situation outside of an experimental set-up. As a result, the two classes of factors, hereditary and environmental, operate simultaneously to produce greater likeness within the ordinary family than is found among individuals chosen at random. The closer the hereditary relationship, moreover, the greater the environmental proximity. Thus parents and children, and brothers and sisters, usually live in the same home; while more distant relatives, such as uncles and nephews, or cousins, come into less frequent contact. Not only are related individuals exposed to common environmental stimulation because of similarity of living conditions, but they also constitute in part each other's environment and may become more alike in some respects through such mutual interaction. It would seem that family groupings offer an excellent example of the operation of environmental influences in the development of behavorial similarities.

Curiously enough, however, family resemblances are often attributed unquestioningly to the operation of heredity. The child is described as having his father's business acumen, his aunt's musical talent, "taking after" his grandfather in obstinacy, and perhaps inheriting a keen sense of humor from an Irish grandmother on his father's side! The successful son of an eminent family attributes his accomplishments to the fact that he is well-born. A lecturer's vigor and zeal are explained by his coming from pioneer stock. A boy's ingenuity with mechanical toys is regarded as only natural when one finds that he is descended from a "long line" of boatbuilders and

inventors.[1] Nor is this type of interpretation limited to popular slipshod thinking and everyday conversation. Many otherwise accurate and well-conducted scientific investigations on family resemblances commit the same logical fallacy in their interpretations.

The two major methods employed in the study of family similarities and differences are family history, or *pedigree studies,* and *correlation.* The former method has been employed chiefly by geneticists. Genealogies are traced and detailed pedigree charts drawn up for families outstanding either for their deficiencies or for their talents. The correlation studies usually deal with the scores of relatives on standardized tests. Parents and children, siblings, and twins have been compared by this method. The correlation coefficient [2] furnishes a convenient numerical index of the degree of correspondence between the scores of any such groups.

It is of course impossible to determine directly by either of these methods what is the relative contribution of hereditary or environmental factors in producing the obtained similarities. Both methods are at best descriptive and serve only to discover more or less objectively the degree of familial resemblance present under existing living conditions. Only an experimental approach could yield a conclusive solution to this problem. If a child of known parentage were isolated from its family immediately after birth and brought up under rigidly controlled conditions, many of the questions on heredity and environment might be answered. In such an experiment, it would also be necessary to exert some control over prenatal environment, as by proper care and diet of the mother. For obvious reasons, such experiments have not been feasible with human subjects. An approximation to this set-up is, however, afforded by the study of foster children. The earlier the child is adopted, the more nearly does this situation resemble the experimental situation described above. A favorable opportunity for the analysis of hereditary and environmental factors is also furnished by identical twins who have been reared apart from an early age, although the number of such cases is necessarily small.

[1] A collection of rather amusing excerpts from biographies of eminent persons, illustrating the common tendency to look for ancestral origins of the individual's talents and defects, is to be found in Tozzer (37).

[2] Cf. Chapter 2 for a general explanation of correlation coefficients. In the present application, the correlation coefficient is used to measure the degree of relationship between the scores of related individuals on the same test, rather than between the scores made by the same individuals on different tests.

Because of their more direct bearing upon the heredity-environment problem, all studies on *twins* and on *foster children* have been reserved for a detailed treatment in the next chapter. The present chapter will deal exclusively with the more common and general sort of family relationships, including parents and children, siblings, and more remote relatives or ancestors.

THE STUDY OF FAMILY PEDIGREES

The tracing of human family pedigrees with reference to some specific and easily identifiable characteristic may reveal valuable data on hereditary factors. The method has proved especially productive in the study of simple physical abnormalities, such as albinism, the presence of extra fingers, webbed fingers, clubfoot, and a number of other rarer and more serious malformations or pathological conditions. Certain simple behavior characteristics may also lend themselves to analysis by these methods. The application of pedigree analyses to more complex behavior data, however, usually meets with well-nigh insurmountable difficulties. Consequently such use as has been made of these methods in the analysis of complex behavior data is on the whole open to serious question. Unwarranted inferences and overgeneralizations abound in these studies.

The identification of hereditary factors from human family histories involves two major steps: *inspection of pedigrees* and *"gene frequency analysis"* (cf. 21, 33). First, a number of family pedigrees in the charcteristic under observation are assembled. From an examination of each of these pedigrees, hypotheses are set up regarding the probable hereditary basis of the particular characteristic. As these hypotheses are checked against other pedigrees, some can readily be discarded, while one may be consistent with all the observed pedigrees and is tentatively accepted. The testing of this tentative hypothesis in representative samples of the general population constitutes the second step, or gene frequency analysis.

In animal studies, this is the stage at which selective breeding and cross-breeding would be carried out as a direct test of the chosen hypothesis. Since this is not feasible in human studies, the procedure is to compare the frequency of different phenotypes in the general population with the frequency expected on the basis of the chosen hypothesis. "Phenotypes" refer to the observably different ways in

which the characteristic in question is manifested in different individuals. For example, in the case of a characteristic determined by a single pair of dominant-recessive factors, an individual may have received two dominant factors from his two parents, or one dominant and one recessive, or two recessives. Those receiving the dominant-recessive combination, however, manifest the dominant characteristic. Consequently, only two phenotypes are found in this characteristic, in contrast to the three different "genotypes" to which they correspond. If the frequencies of dominant and recessive genes for this characteristic were identical in the general population, then the two phenotypes would occur in the well-known Mendelian ratio of 3:1. Ordinarily, however, the two genes will not be equally common, and the simple 3:1 ratio will not hold. Nevertheless, under these circumstances *certain constant relationships* between the frequencies of different phenotypes will be found.[3] It is these relationships that are employed in the gene frequency analyses. Such relationships can be derived for various types of hereditary mechanisms, such as dominant-recessive, blending, and sex-linked characteristics, as well as for characteristics depending upon more than one pair of genes.

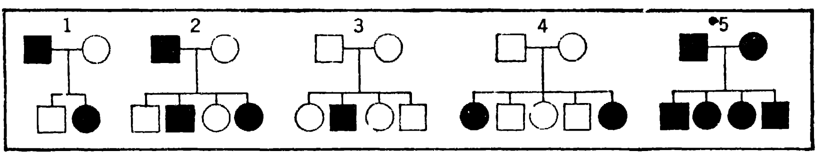


Fig. 63. Selected Pedigrees of Taste Deficiency in Man. (From Snyder, 33, p. 416.)

The use of family pedigree techniques is well illustrated by the study of taste deficiency in man. Quite accidentally it was discovered that some persons report no taste from the crystals of a certain chemical, phenyl-thio-carbamide (P.T.C.). To most people, these crystals are very bitter. It was soon suspected that this difference might have a genetic basis and investigation of its possible hereditary transmission was begun. In Figure 63 are reproduced five family pedigrees for this characteristic, selected from several thousand which were examined (cf. 4, 33). The first pedigree leaves the way open for many different types of hereditary determination, but different hypoth-

[3] All these relationships are derived from the statistics of probability as applied to the chance pairing of genes.

eses can be eliminated successively as each additional family is considered. For example, family No. 3 definitely shows that this taste deficiency cannot be attributed to a dominant factor, since the deficiency appeared in a child both of whose parents were free of it. The suggestion that a single recessive may be involved is borne out by family No. 5, where both parents show the deficiency. In this family, as expected, *all* the offspring are deficient. Other hypotheses, regarding the possibility of sex-linked and sex-influenced factors, can be ruled out by an inspection of the other two pedigrees.

Following the tentative acceptance of the hypothesis of a single pair of dominant-recessive genes, a gene frequency analysis was conducted on a random sample of 800 families. These families included some in which both parents were normal tasters, others in which both were deficient, and still others with one normal and one deficient parent. The proportion of tasters and non-tasters among the offspring in each of the three types of families, as well as the proportion of tasters and non-tasters in the general population, constitutes the basic data for the gene frequency analysis. If the chosen hypothesis holds, certain relationships are expected among these various proportions. In Table 13 are shown the observed and expected per cents of non-tasters among the offspring of each type of family.[4] If taste deficiency depends upon a single recessive factor, all offspring of two non-taster parents should be non-tasters. That the obtained per cent is 97.76 rather than 100, owing to the presence of five tasters in this category, need not be regarded as evidence against the hypothesis. The investigators (4) suggest a number of possible reasons to account for these exceptions. the subjective nature of the taste experience may have led to incorrect diagnoses; parentage may have been incorrectly determined because of unsuspected adoption or illegitimacy; mutations or unknown factors of a hereditary or environmental nature may have affected the operation of the recessive gene. In the other two types of families, it will be noted that the observed and expected percentages agree closely and thus confirm the hypothesis of a single recessive gene.[5]

[4] For an explanation of the computation of the expected percentages, cf. 4; 32; and 33, Ch. 29.

[5] The statistically trained reader will note that the differences in both groups are smaller than their respective standard errors and are therefore well within the range of variation to be expected from sampling errors. Neither of the differences is thus statistically significant. (The concept of statistically significant difference will be explained in Chapter 18.)

TABLE 13 *Gene Frequency Analysis of Taste Deficiency for P.T.C.*

(From Cotterman and Snyder, 4, p. 514)

Type of Family	*Number of Families*	*Total Number of Offspring*	*Per Cent of Non-Tasting Offspring*		
			Observed	*Expected*	*Difference*
Both parents non-tasters	86	223	97.76	100.00	2.24
One parent taster, the other non-taster	289	761	36.53	35.32	1.21 ± 1.76*
Both parents tasters	425	1159	12.28	12.47	0.19 ± 1.02*

* Standard error of the difference.

When either the pedigrees or the observed frequencies are not consistent with *any* unit-factor hypothesis, other hypotheses are set up in terms of two or more pairs of factors. For example, with two pairs of dominant-recessive factors, four phenotypes will be found. Even more phenotypes result when there is a lack of dominance in one or more pairs of factors. Under these conditions, the frequency patterns become more complicated, but they are still predictable and therefore amenable to testing. When the number of hereditary factors involved is very large, however, an almost infinite number of quantitative gradations is found, rather than distinct phenotypes. In such cases, the frequency distribution approaches the normal curve.

The *multiplicity of hereditary factors* contributing to most behavior functions is one of the obstacles encountered in the application of family pedigree methods to human behavior data. If the observed frequencies follow the normal curve, little can be deduced beyond the operation of a very large number of factors. Moreover, the same frequency distribution could result from the combined effect of the innumerable environmental influences to which the developing individual is exposed. Such results certainly do not permit the same clear-cut interpretation which is possible when simpler genetic ratios are involved.

A second disturbing factor in such analyses is the indisputable operation of *assortative mating* in human marriages. Gene frequency

analyses are based upon the assumption of random mating. This assumption is probably justified, on the whole, with regard to such characteristics as the taste deficiency described above, since most individuals are not even aware of this deficiency in either themselves or their associates. Moreover, this deficiency appears not to be correlated with other characteristics which might enter into assortative mating, such as general appearance, physique, intellectual level, socio-economic level, or national, racial, or geographical background. However, most behavior characteristics—and many physical characteristics—either play a direct part in assortative mating or enter indirectly through their association with socio-economic level, geographical distribution, and the like. Individuals tend to marry within their own groups, economically, nationally, geographically, and intellectually. Husband-wife correlations in intelligence tests, for example, are generally in the neighborhood of .50, and in physical traits they cluster around .25 (16). In personality characteristics, the correlations vary widely, as would be expected. In the more purely emotional characteristics, such as emotional stability and social dominance, the correlations are relatively low and sometimes negative, averaging about .14 (5). On tests of attitudes and values, the correlations range from the .20's to the .70's and average about .59 (5). To be sure, such marital correlations may result in part from the common experiences and mutual influence of the spouses after marriage. It is doubtful, however, whether such influences can account for a large part of the observed correlations, especially since many of the subjects of these studies had not been married long. Most of the correlation can thus be safely attributed to assortative mating, or the tendency for similar individuals to marry.

A third ever-present difficulty in the genetic analysis of human behavior data is the influence of *environmental factors*. The testing of genetic hypotheses implies either a constant influence of environment or random environmental variation. In actual fact, however, environmental differences among individuals are not random, but tend to go hand in hand with hereditary differences. Thus the child of physically defective or feebleminded parents is also more likely to have low socio-economic level, poor physical care, and inferior education than is the child of intellectually and physically superior parents.

A further difficulty is presented by the likelihood of *inaccurate and incorrect diagnosis,* especially when information is sought regarding

individuals who have been dead for many years. The data collected in retrospect on feebleminded ancestors, for example, are often based upon reports by untrained persons or upon inadequate records. Another difficulty, in the reverse direction, is encountered when gathering information on *characteristics which are not manifested until late in life*. For example, certain psychoses usually develop among older persons. Information on these conditions cannot, therefore, be obtained while the subjects are still young. Moreover, some individuals die before reaching the age when such conditions might have developed.

THE FAMILIES OF EMINENT MEN

It should be apparent that the mere recurrence of a characteristic within a family pedigree proves nothing regarding its hereditary determination. The proper genetic study of family pedigrees, as shown in the preceding section, involves much more than the simple fact of family resemblances. Nevertheless, because most human behavior characteristics do not lend themselves to the precise methods of analysis outlined above, many widely quoted studies of family pedigree provide little or no information beyond the greater similarity of behavior among related than among unrelated persons. This type of familial investigation was launched by the publication, in 1869, of Sir Francis Galton's *Hereditary Genius*.

Galton's approach was distinctly hereditarian, as illustrated by the following summary of the aim of his investigation: "I propose to show in this book that a man's natural abilities are derived by inheritance, under exactly the same limitations as are the form and physical features of the whole organic world" (10, p. 1). Data were collected on 997 eminent men in 300 families. In order to facilitate the tracing of family histories and the location of descendants and other relatives, the study was limited to eminent men who were either English or well known in England. The information was obtained from biographical collections or through direct inquiry among relatives and acquaintances of the men themselves. Galton defined as follows the degree of eminence necessary for inclusion in his survey: "When I speak of an eminent man, I mean one who has achieved a position that is attained by only 250 persons in each million of men, or by

one person in each 4000" (10, p. 9). The classes of men in Galton's survey comprised English judges,[6] statesmen, commanders, literary men, scientists, poets, artists (musicians and painters), and Protestant divines, the last including men who had achieved fame through some phase of religious activity, such as theological scholars, administrators, religious leaders, martyrs, preachers.

Within each family, the most eminent man was taken as a point of reference, and all kinships were expressed in relation to him. Following the name of each of these men, Galton appended a list of famous relatives together with the major field in which each had achieved distinction. Whenever more complete information was available, these data were presented in the form of a family pedigree chart. As a final summary of his findings, Galton computed the percentage of eminent men in each degree of kinship to the most eminent man of the family, the latter still serving as the point of reference. These percentages are given in Table 14 for each class of "eminence" separately, as well as for all classes combined. It should be noted that

TABLE 14 *Percentage of Eminent Relatives of Men in Each Class*

(From Galton, 10, p. 308)

Nature of Kinship *	*Judges*	*States-men*	*Com-manders*	*Literary*	*Scientific*	*Poets*	*Artists*	*Divines*	*All Classes*
Father	26	33	47	48	26	20	32	28	*31*
Brother	35	39	50	42	47	40	50	36	*41*
Son	36	49	31	51	60	45	89	40	*48*
Grandfather	15	28	16	24	14	5	7	20	*17*
Uncle	18	18	8	24	16	5	14	40	*18*
Nephew	19	18	35	24	23	50	18	4	*22*
Grandson	19	10	12	9	14	5	18	16	*14*
Great-grandfather	2	8	8	3	0	0	0	4	*3*
Great-uncle	4	5	8	6	5	5	7	4	*5*
First cousin	11	21	20	18	16	0	1	8	*13*
Great-nephew	17	5	8	6	16	10	0	0	*10*
Great-grandson	6	0	0	3	7	0	0	0	*3*
All more remote	14	37	44	12	23	5	18	16	*31*

* No female relatives are included in these summary figures, although the names and achievements of such relatives are given in the specific family histories.

[6] The only category limited exclusively to England.

the eminent relatives within any class have not necessarily achieved distinction in that particular area; thus the famous kinsmen of a statesman may include scientists, artists, divines, etc. The classification is based solely on the field of activity of the "most eminent" man in the family, around whom the data are organized.

These figures suggest quite strongly that eminence tends to run in families. Not only are the percentages much greater than is expected by chance and fairly consistent from class to class, but they also show a definite decrease in the frequency of eminent relatives as the degree of relationship becomes more remote. It is quite a different matter, however, to conclude that genius is inherited. Galton, to be sure, recognized the difficulties in the way of such a conclusion and attempted a systematic analysis of them. To the question of whether reputation is a fair test of ability, he answers in the affirmative. He argues that reputation or eminence, as the criterion is employed in his survey, is "the opinion of contemporaries, revised by posterity—the favorable result of a critical analysis of each man's character, by many biographers" (10, p. 33), and hence is not an accidental rise to short-lived notoriety. Natural ability he defines quite circularly as "those qualities of intellect and disposition, which urge and qualify a man to perform acts that lead to reputation" (10, p. 33).

Although admitting the influence of training, surroundings, and opportunities, Galton minimizes the part which they play in the attainment of eminence. He constantly holds up to the reader the heroic picture of genius triumphing over obstacles. By definition, genius means to him "a nature which, when left to itself, will, urged by an inherent stimulus, climb the path that leads to eminence, and has strength to reach the summit—one which, if hindered or thwarted, will fret and strive until the hindrance is overcome, and it is again free to follow its labour-loving instinct" (10, pp. 33–34). He concludes that "It is almost a contradiction in terms, to doubt that such men will generally become eminent," and adds that "there is plenty of evidence in this volume to show that few have won high reputations without possessing these peculiar gifts" (10, p. 34). This is true enough, but it remains to be proved that such "gifts" as the impulse to climb, the strength to reach the summit, and the love of labor are themselves independent of environment. Unfortunately, the optimistic picture painted by Galton is not borne out by observations of

everyday life; and in the absence of empirical proof, it is impossible to accept Galton's interpretations of his findings.[7]

DEGENERATE FAMILIES

The family history method has also been widely employed in the effort to analyze the causes of intellectual defect, crime, pauperism, and similar conditions. By this method, a number of families have been discovered which present an overwhelming array of socially inadequate persons over several generations. The same general techniques are used in tracing the history of these families as in the study of eminent groups. Living relatives or descendants are visited and observed, residents of the vicinity are interviewed, and certificates of marriage and birth and similar public records are examined whenever available. These families are usually found in rural districts in many parts of the country, often inhabiting the same crude huts built by their ancestors many generations ago. They interbreed extensively, are quite prolific, and eventually come to constitute their own community, avoided and ridiculed by their neighbors.

The earliest published pedigree of such a "degenerate" family is that of the "Jukes,"[8] described by Dugdale (7) in 1877 and subsequently traced up to 1915 by Estabrook (8). This family first attracted official notice in the course of a prison survey in New York State in 1874. Six persons, all of whom were blood relations, were found in prison in one county. This finding initiated a thorough search for other relatives living in the county and finally led to an extensive family history, which covered seven generations and included 540 individuals related by blood and 169 related by marriage or cohabitation. The total cost of this family to the state through pauperism, crime, vice, disease, and similar conditions was estimated as one and one-half million dollars within 75 years.

The original Jukes were five sisters or half-sisters whose progeny, legitimate and illegitimate, have been traced for five generations. Two of these sisters married two sons of "Max," a descendant of the early Dutch settlers, who lived as a backwoodsman and is described as a

[7] Other studies on eminent families will be found in Chapter 17 on *Genius*. Galton's study is here reported only as an example of this application of the family history method.

[8] All the names in these histories are, of course, fictitious, but they have become well known in the psychological literature.

"hunter and fisher, a hard drinker, jolly and companionable, averse to steady toil" (7, p. 14). This man was born in New York State between 1720 and 1740. The genealogy of the Jukes is usually begun with Max, although it is the progeny of the five sisters who have been traced and are shown in the pedigree charts.

It is interesting to note that, despite the fact that family histories are usually cited as examples of hereditary characteristics, Dugdale seemed to be fully cognizant of the influence of environment, as is shown by the following conclusion: "From the above considerations the logical induction seems to be, that environment is the controlling factor in determining careers . . . the permanence of ancestral types is only another demonstration of the fixity of the environment within limits which necessitate the development of typal characteristics" (7, p. 66). And again, in a final summing up of his findings he calls attention to the fact that "In the 'Jukes' it was shown that heredity depends upon the permanence of the environment, and that a change in the environment may produce an entire change in the career, which, in the course of greater or less length of time, according to varying circumstances, will produce an actual change in the character of the individual" (7, p. 113).[9]

The "Kallikak" family of New Jersey, described by Goddard (12), is particularly interesting since it consists of two branches, one normal, the other degenerate. The history of this family has been traced to the days of the American Revolution. "Martin Kallikak," a 21-year-old youth of good family, who had joined one of the many military companies organized at the time, had sexual relations with a feebleminded girl whom he met at a tavern. The illegitimate child of this union, referred to as "Martin Kallikak, Jr.," was the progenitor of the degenerate side of the family. Martin, Sr., at the age of 23 married an intellectually superior woman of his own social level and thereby founded a normal family, many of whose members have achieved distinction. A pedigree chart showing the normal and degenerate lines of the Kallikak family is reproduced in Figure 64. The forbears of Martin, Sr., are shown for three generations, and the two

[9] Winship (40) has contrasted the Jukes with the Edwards, a distinguished family descending from Jonathan Edwards, a highly educated and famous theologian of eighteenth-century America. The comparison is striking, but not very informative, since the two families were entirely independent, of different ancestral stock, and living in very different environments.

branches which he founded are traced through the line of the eldest son to a member of the present generation.

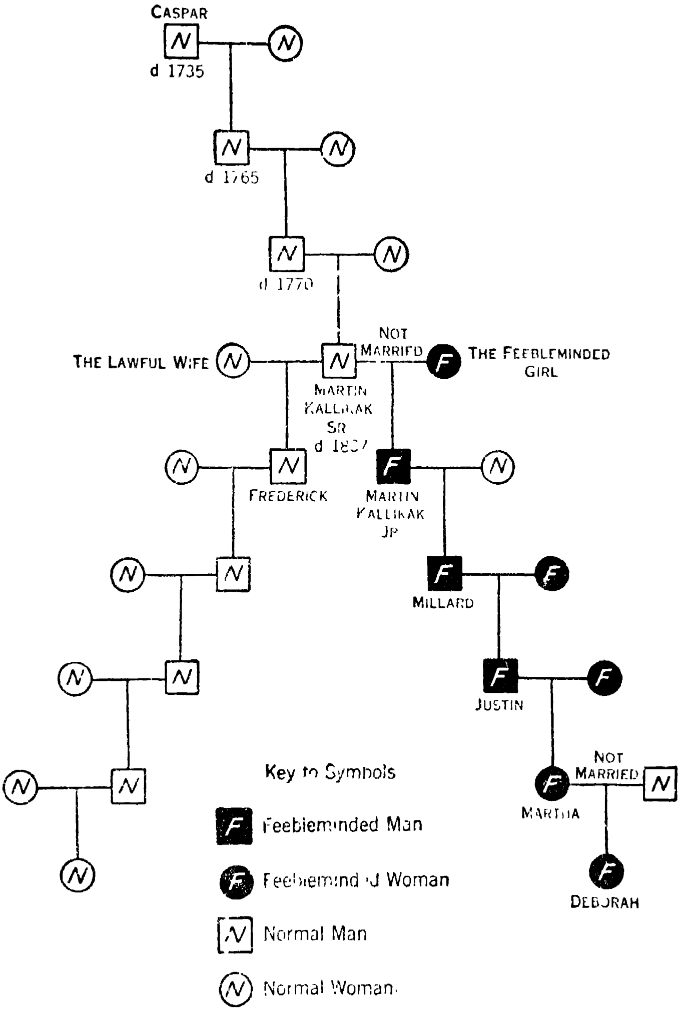


Fig. 64. A Pedigree Chart of the Kallikak Family (From Goddard, 12, p. 36.)

In evaluating the findings on the Kallikak family, Goddard constantly emphasizes the role of heredity. Having laid great stress upon the fact that the two groups were branches of the same family,

furnishing, "as it were, a natural experiment with a normal branch with which to compare our defective side," he states that "from this comparison, the conclusion is inevitable that all this degeneracy has come as the result of the defective mentality and bad blood having been brought into the normal family of good blood" (12, pp. 68–69). It seems rather curious that the common descent of the two branches from Martin Kallikak should be regarded as strengthening a hereditary interpretation of the differences between them. The environments of the two groups were not in any way equated by this common ancestry. In fact, it is evident that the members of the two branches were reared under widely differing conditions. A more crucial test would have been available if the legitimate offspring of Martin and his well-born wife had been exchanged at birth with those of the feebleminded woman. It would then have been very illuminating to ascertain the relative percentage of feeblemindedness and other defects in the "normal" and "degenerate" stock. The practical obstacles in the way of such a procedure in no way excuse faulty conclusions drawn from an inadequately controlled situation.

Leading geneticists have been critical of the Kallikak study since it first appeared. Goddard's assumption that feeblemindedness is transmitted by a single recessive gene seems indeed a gross over-simplification in the light of present knowledge of heredity. More specifically, Goddard maintained that Martin Kallikak, Sr., must have had a recessive gene for feeblemindedness, which would account for the recurrence of feebleminded descendants from his union with the feebleminded girl. If this had been the case, however, the complete absence of feebleminded offspring in the "good branch" of the family would be difficult to explain. Criticism has also been directed against much of the data which forms the basis of this family history (e.g., 29, 30). For example, the only evidence for the paternity of Martin Kallikak, Jr., is based on the original report of the feebleminded tavern girl. The IQ's of long-dead Kallikaks were often estimated on the basis of the reminiscences of elderly neighbors. Moreover, in tracing the family pedigrees the rather questionable assumption was made that such conditions as pauperism, crime, immorality, and epilepsy are all manifestations of the same recessive gene which produces feeblemindedness.

Many equally "degenerate" families have been subsequently in-

vestigated by psychologists, sociologists, or eugenicists. The research staff of the Eugenics Record Office [10] conducted many such surveys as one phase of its regular work. Among the groups thus studied were the Hill Folk, the Nam family, and the W family of Indiana, all presenting the same picture of degeneracy, mental defect, disease, and social incompetence through successive generations. Surveys of eminent families have likewise been sponsored by the Eugenics Record Office. Specific lines of achievement, such as scholarly pursuits or boat designing, have been traced from generation to generation in the attempt to show that such talents are transmitted through heredity. Although offering much interesting material, such studies cannot yield any data on the heredity-environment question; the opportunities for environmental transmission of such family qualities are too obvious to overlook or dismiss.[11]

Among the most recent applications of the family pedigree method, greater care has been exercised to insure the accuracy of the original data (cf., e.g., 15, 20, 38). Diagnoses of feeblemindness among relatives in earlier generations, for example, are examined more critically and accepted only when verified by institution records or commitment papers. Nevertheless, the large majority of these studies are still subject to the fallacy of regarding mere recurrence of a characteristic within the family as proof of heredity. Only rarely is any attempt made to suggest a specific hypothesis of hereditary transmission (cf., e.g., 1). Nothing even remotely resembling the type of evidence described in the opening section of the present chapter is provided. Most investigators are apparently interested only in showing that the condition "runs in families." To regard these studies of feebleminded or of eminent families as applications of the genetic methods of pedigree analysis can only lead to confusion.

[10] Eugenics Record Office, Cold Spring Harbor, Long Island, N. Y. Cf., e.g., references (6) and (9).

[11] A typical example of the misuse of data on family resemblances is to be found in an article by Rife and Snyder (26). Thirty-three contemporary case histories of "idiots savants" in American institutions are described and, on the basis of certain familial resemblances, are offered as a "refutation" of environmental determination of mental development. The very fact that these feebleminded subjects exhibited some special talent seems also to be regarded, by a peculiar logic, as evidence for heredity. Among the cases cited is that of a low-grade idiot who could spin objects rapidly with either hand, balancing them on his index finger. Both of his parents were vaudeville actors! This case was presented in all seriousness as "evidence" for the inheritance of special talents.

PARENT-CHILD RESEMBLANCE

The use of the correlation technique, although more precise, does not eliminate the essential difficulty inherent in all family comparisons, namely, the confusion of hereditary and environmental contributions. Pearson (23) was among the first to apply correlation analysis to parent-child resemblances. Continuing a line of research initiated by Galton (11), he collected measures on parents and offspring in physical traits such as stature, arm span, and forearm length. The parent-child correlations in these traits averaged about .52. The similarity of this correlation to those obtained for bodily characteristics of many animal forms led Pearson and others to suggest that this figure indicates the contribution of hereditary factors to the development of physical traits. Family resemblance in such traits is probably attributable in large part to heredity, although the influence of similar environment, especially in the prenatal stage, cannot be overlooked.

More recently, scores obtained by parents and children on standardized psychological tests have been correlated. In the most extensive of these studies, Conrad and Jones (3) administered intelligence tests to 269 family groups, including 977 persons between the ages of 3 and 60. All subjects were native-born, spoke only English at home, and lived in rural districts of New England. Socio-economic differences within this sampling were small. The younger subjects were tested with the Stanford-Binet, the older with the Army Alpha Intelligence Examination. For the entire sampling, the total parent-child correlation obtained with these tests was .49. No consistent or significant difference was found between mother-child and father-child correlations, nor did the correlation of sons or daughters with their like-sex parent differ from the correlation with their unlike-sex parent. It might be argued that if environment is important in producing these familial resemblances, then children should resemble their mother more closely than their father. It is true that the mother generally has closer contact with the children than does the father, but it may also be noted that the father's intellectual level probably determines the socio-economic level of the home more than does that of the mother. Conrad and Jones demonstrate statistically that the obtained correlation of .49 is consistent with an hereditary interpretation of parent-child resemblances in intelligence, after allowance is made for assortative mating. They recognize, however, that the results

are equally consistent with a purely environmental hypothesis, or with a combination of hereditary and environmental influences.

In generalizing from the specific correlation found in this study, two further facts should be borne in mind. First, the parent-child correlation varies with the *nature of the test*. For most intelligence tests, which are a composite of many tasks of a predominantly verbal nature, the correlation of about .50 is probably typical. The correlation on more homogeneous and simpler tasks will, in general be lower. Non-verbal functions, moreover, tend to give lower correlations than the more highly verbal (39). Performance in verbal functions is probably more dependent upon differences in previous experience and home background, a fact which may account for the closer family resemblances on verbal tests.

A second factor which affects familial correlations is the degree of *homogeneity of home background* within the group. In the Conrad-Jones study, it will be recalled, the sampling was particularly homogeneous. The authors call attention to this fact, pointing out that the apparent influence of a common home environment within each family is minimized when the differences from home to home are slight. Thus the correlations between parents and children might be much higher if a wider range of homes were sampled.

It should also be noted that the parent-child correlation of approximately .50 in intelligence test scores is not found until the child is about 5 years of age (28). The correlation is considerably lower at earlier ages and approaches zero in infancy. It will be recalled that a similar lack of correlation was found between the individual's score in infancy and his own later performance. The two findings probably have a similar explanation. A principal factor in such an explanation is undoubtedly the difference in behavior functions tested among preschool children and among older children or adults.

Parent-child correlations in personality test scores also tend to be positive and significant, although running lower than intelligence test correlations (5). The correlations vary widely with the particular aspect of personality under consideration. On the whole, the degree of parent-child resemblance indicated by the available data appears to be lower for emotional characteristics, such as introversion, dominance, or neuroticism, and higher for attitudes. In fact, the average parent-child correlation on most attitude scales is approximately as high as on intelligence tests. In connection with the relatively low

correlations on tests of emotional characteristics, it is interesting to consider the possible effects of parental personality upon the development of the child's personality. It is likely, for example, that excessive dominance in a parent may foster the opposite type of reaction in the child. The effects of parent-child interaction probably differ widely with the degree of the personality characteristic manifested, as well as with many other attendant circumstances.

THE COMPARISON OF SIBLINGS

The study of siblings, especially when both are in school, does not present the practical difficulties met in testing parents. Consequently, investigations on the resemblance of siblings are more plentiful, over a dozen studies on adequately large samplings being on record. In the previously cited study by Conrad and Jones (3) on familial resemblance, a total of 644 individual siblings in 225 families were tested. The correlation was identical with that found for parents and children in the same study, viz., .49. That the sibling correlation on most intelligence tests is in the neighborhood of .50 has been repeatedly confirmed. The correlation between 384 pairs of siblings tested during the standardization of the revised Stanford-Binet Scale (19) was found to be .53. The same correlation (.534) was obtained with about 650 pairs of siblings tested in Scotland (27). The latter group was especially free from limitations of sampling, since it represented all siblings located during a project in which every child born in a given community within certain dates was tested.

Under various conditions, the sibling correlation in mental test scores may drop as low as .30 or rise to nearly .70 (14). *Heterogeneity* of the samplings tested is undoubtedly a factor in some of these differences. Correlations as a whole tend to be higher in more heterogeneous groups in which the scores range more widely. Among college students, who represent a much more homogeneous group than the general population, the sibling correlation in intelligence test scores is closer to .40 than to .50 (34, 35). When siblings attending a single school are tested, however, the influence of the *common school environment*, together with *selective factors*, may exert the opposite effect upon the sibling correlation. Thus if one member of a sibling pair is attending college and the other is not, such a pair would automatically be excluded from the study. But

these are the very pairs likely to show the largest differences in test performance. Their omission would therefore raise the apparent correlation between siblings. In a high school sampling, for example, in which selection and a common school environment probably had more effect than the slight increase in homogeneity, the sibling correlation on an intelligence test was .60 (36). As in the case of parent-child correlations, the *nature of the test* also affects the size of sibling correlations, the more verbal type of tests tending in general to yield higher correlations (39).

Sibling correlations show no consistent trend either to rise or drop with *age*, when the same intelligence test is used throughout (19). It is of course true that the older the subjects, the longer will environmental factors have operated upon them. But whether such factors exert a leveling or a differentiating influence upon the development of siblings within any one family obviously depends upon whether the environments of the siblings have remained similar or diverged with age. If, for example, one sibling goes away to boarding school at age 10, while the other remains at home, it would hardly be reasonable to expect environment to make them more alike with age just because they are members of the same family.

The amount of *age discrepancy* between siblings also appears to have little or no effect upon sibling correlations in intelligence test scores (19, 25). For an interpretation of such findings, much more information is needed regarding the social interaction of siblings with each other and with their parents. A preliminary effort to investigate such social factors, especially as they affect the intelligence test performance of older and younger siblings, is illustrated by an intensive follow-up study of 39 pairs of siblings, conducted as a part of the Fels Growth Study (17). All the sibling pairs consisted of a first-born and a second-born child. The children, ranging in age from 30 months to 12 years, were tested at regular intervals with alternate forms of the Stanford-Binet. With such data, it was possible to compare the performance of first- and second-born siblings in each family on tests administered *at the same age*. The two siblings in each pair were thus compared on the same test items. Significant differences in the frequency with which first-born and second-born siblings passed certain Stanford-Binet items were found. In general, the first-born siblings tended to excel on relatively abstract, verbal items, while the second-born were superior on a larger number of items, and espe-

cially on items involving realistic, concrete tasks. The type of intellectual stimulation received by the first-born child, who is more likely to have adult companionship, is suggested as one possible factor to account for these differences.

The comparison of test correlations between *like-sex* and *unlike-sex* siblings shows no consistent differences (3). One might expect a closer resemblance between like-sex siblings because of greater similarity of experience. The interaction and mutual influence of children within the family may be such, however, as to counteract the similarities in the environments of like-sex siblings. When possible sibling rivalries and similar motivational factors are considered, it is apparent that no simple relationship between the development of like-sex and unlike-sex siblings can be predicted.

As is true of parent-child correlations, sibling correlations on *personality tests* are lower, in general, than on intelligence tests. When ratings are employed, as in a pioneer study by Pearson (22), the sibling correlations will be spuriously high because of the rater's tendency to rate two members of the same family alike. On groups of 500 or more siblings, Pearson found sibling correlations in the .50's and .60's in such traits as "vivacity" and "self-assertiveness." In contrast to these results with ratings, test scores have yielded correlations of about .15 in emotional adjustment, introversion, and similar characteristics (5, 24). On attitude scales, the sibling correlations are higher, clustering between .30 and .40 (5). In their extensive study of character traits among school children, May and Hartshorne (18) compared the performance of 734 pairs of siblings. The sibling correlations on tests of honesty ranged from .21 to .44; in persistence and inhibition, the correlations ranged from .14 to .46, and in service and self-sacrifice, from .05 to .40 (5, 18).

What are the implications of sibling studies for the problem of heredity and environment? Some have pointed out that the intelligence test correlation of approximately .50 found between siblings in the general population closely resembles the correlation to be expected for a characteristic determined by multiple factor heredity (27). Nevertheless, the fact remains that the obtained correlation lends itself with equal facility to other interpretations, and no one hypothesis can therefore be accepted solely on the basis of such a correlation. Attempts have also been made to compare the sibling

correlations in psychological and in structural characteristics, in an effort to disentangle the relative contributions of heredity and environment (22, 36). It has been argued, for example, that since the sibling correlation in such traits as height and intelligence is very similar, and since height can be little influenced by environment, then intelligence must be equally independent of environment. This argument begins by assuming that psychological and physical traits are influenced to an equal degree by heredity. Any influence of environment upon psychological traits would then be superimposed upon this common hereditary influence and would be expected to raise the correlation for psychological traits. Such an argument obviously begs the question.

In this connection may also be considered the implications of sibling correlations in animal studies. In an investigation of maze learning in white rats (2), for example, a sibling correlation of .31 was found in the error scores.[12] Since all the rats were living under fairly uniform conditions, this sibling correlation obviously cannot be attributed to environmental differences among the "rat families," but rather indicates the influence of hereditary structural factors upon maze learning. That such factors do operate in maze learning was, of course, indicated in the selective breeding experiments previously discussed (cf. Ch. 5). Does this clearly non-environmental sibling correlation in the rat experiments suggest that the sibling correlations in the human studies are likewise determined principally by hereditary factors? Not at all. There is no basis for supposing that the same or similar structural factors which operate in a motor learning situation in white rats also operate in the behavior sampled by human intelligence tests. We cannot generalize from one situation to the other, any more than we could generalize from studies on sensori-motor learning in infants to the learning of calculus by college students, in our earlier discussion of maturation and learning (Ch. 5).

An interesting illustration of the fact that similar correlations may have very different origins is furnished by an investigation on Louisiana public school children in grades 5 to 11 (31). Having located 203 pairs of siblings in these grades, the investigator paired each

[12] Because some of the litters classified as independent may have actually been half-siblings, the authors suggest that their group may have been atypically homogenous and the obtained correlation consequently too low. It is therefore likely that such a sibling correlation should be somewhat higher than .31.

child with his own sibling and also paired him with an unrelated child of the same age, of similar socio-economic background, and attending the same school. The intelligence test scores of these unrelated pairs of children correlated .35, only slightly lower than the correlation found between siblings in the same study. Had the home backgrounds of the unrelated children been paired off more precisely and on the basis of a larger number of characteristics, the correlation between their intelligence test scores might have been even higher.

In conclusion, the study of family resemblances in complex intellectual and emotional characteristics, whether by correlation or by other techniques of pedigree analysis, does not furnish any unambiguous clues to the origin of such resemblances. The results do suggest the complexity of factors which operate within the usual family milieu. Despite the superficial uniformity of environment, some of the interactions among individuals in the family group may make for similarity of psychological development, while others may produce progressively divergent trends of behavior. These considerations have prepared the way for an understanding of some of the findings on twins and foster children, to be discussed in the following chapter.

CHAPTER

11

Twins and Foster Children

CERTAIN SPECIAL FAMILY RELATIONSHIPS have been singled out by investigators as offering a more direct opportunity to disentangle the contributions of heredit.ry and environmental factors. Chief among the groups studied for this purpose are *twins* and *foster children.* Attention has also centered upon children reared in *institutional environmens,* such as orphanages. Twins and foster children fall at opposite poles in respect to hereditary similarity. In the case of *identical twins,* heredity is completely alike for the two individuals, since they develop from a single fertilized ovum and thus have identical sets of genes. At the other extreme, foster children are reared in a family unit with which they have no hereditary connection whatsoever. It follows that any *difference* between identical twins must result from the operation of environmental factors. Conversely, *similarities* between foster children and their foster parents or foster siblings suggest the influence of the common home environment.

The study of *fraternal,* or non-identical, twins also provides a promising approach to this general problem. Such twins are no more alike than ordinary siblings in respect to heredity. They have, however, been exposed to a similar prenatal environment, since they developed and were born at the same time. Being of identical age, they are also exposed to more nearly similar stimulation throughout childhood than are ordinary siblings. They would thus seem to offer a sort of "hereditary control" in the analysis of the sibling resemblances which have ordinarily been observed. Similarly, identical twins who have lived apart from early infancy may be regarded as a "hereditary control" for identical twins reared together in the usual way. Within each of

these two comparisons, the degree of hereditary resemblance is the same. Any differences in the degree of behavioral similarity can thus be traced to environmental factors.

Children who are brought up in orphanages or similar institutions are in a more uniform environment than those living either with their own parents or in foster homes. For this reason, special interest attaches to the resemblances and differences which these children show among themselves; their resemblance to the parents from whom they were separated is likewise of interest. Any relationship between the nature of the institutional program and facilities on the one hand, and the children's behavioral development on the other, is also relevant.

THE STUDY OF TWIN RESEMBLANCE [1]

Beginning with the pioneer study of Galton in 1875 (cf. 14), twin resemblances have served as the nucleus for a number of investigations on heredity and environment. The earlier studies on twin resemblance in mental test performance failed to differentiate between fraternal and identical twins, thus precluding a clear-cut interpretation of their results. All agreed in finding a closer resemblance between twins than between siblings, and greater similarity between like-sex than between unlike-sex twins. The latter must obviously be fraternal, since identical twins are always of like sex. The groups of like-sex twins, on the other hand, undoubtedly included some fraternals along with the identicals.

In more recent investigations, the two types of twins are generally considered separately, and an increasing use is being made of more refined and dependable methods of classification. If the twins are enclosed within a single sac (or chorion) at birth, it is certain that they are identical. Such information, however, is not always available. Moreover, two-sac pairs which are derived from a single fertilized ovum do occasionally occur. This criterion cannot therefore be relied upon exclusively as a means of separating the identicals from the fraternals. The safest procedure is to compare the twins in a fairly large number of physical characteristics. Close similarities

[1] For a non-technical introduction to many of the biological questions regarding twins and other multiple human births, cf. Newman (46). An excellent critical survey of the psychological findings on twins, as well as on foster children and institutional groups, can be found in Woodworth (77, 78).

might occur by chance in two or three such traits, but if the twins are alike in a *combination* of several characteristics, it is well-nigh certain that they are identicals. Among the most dependable criteria are similarities in fingerprints, hand- and footprints, color of hair and eyes (including the detailed pattern of iris pigmentation), form and texture of hair, and shape and arrangement of teeth. Identical twins must likewise belong to the same blood group, and since a large number of such groups have now been identified, this comparison also provides a fairly good index.[2]

Typical results obtained when different types of twins are compared in intelligence test performance are illustrated by a study of over 375 children with the Stanford-Binet (72). The average difference in IQ within each pair of identical twins, like-sex fraternals, and unlike-sex fraternals is given below, together with the average difference among ordinary siblings:

63 pairs of identicals	5.08
39 pairs of like-sex fraternals	7.37
84 pairs of unlike-sex fraternals	8.48
199 siblings	13.14

The degree of twin resemblance may also be expressed in terms of the correlation coefficient. The two types of data—average difference within pairs and correlation between paired scores—are mathematically equivalent and one can be predicted from the other.[3] The reader may find it more convenient to visualize the relationship in terms of one or the other comparison. The correlations between intelligence test scores of identical twins are generally in the .90's, nearly as high as the reliability coefficients of the tests themselves. In other words, the resemblance between identical twins reared in the same home is about as close as that between test and retest scores of the same individual. The correlations between intelligence test scores of fraternal twins fall between those of identical twins and those of siblings. Such correlations are more variable from study to study than almost any other type of familial correlation, ranging from slightly over .50 (cf., e.g., 23) to about .70 (47, 76).

This finding is not surprising when we realize that the identification

[2] It is interesting to note that recent studies in electroencephalography also seem to indicate that one-egg twins manifest identical brain wave patterns (cf. 78).

[3] Cf 48. Mean intra-pair difference = $1.1\sigma\sqrt{1-r}$, in which r is the correlation between pairs and σ is the standard deviation of the scores within each paired group.

and classification of fraternal twins is subject to special *selective factors* which may operate differently in different studies. On the one hand, fraternal twins who are quite dissimiliar in appearance and behavior are more likely to be overlooked in any search for twins. Such pairs tend more often to be regarded as ordinary siblings in a cursory survey of, for example, the children in a particular school. This selective factor would lead to an overestimation of the correlation between fraternals, since the less similar pairs are omitted. On the other hand, when the classification of twin pairs into fraternals and identicals is somewhat superficial, those fraternals who are most nearly alike in physical and behavioral characteristics are likely to be mistaken for identicals. This will have the effect of reducing the fraternal correlation, because the more similar pairs are now eliminated from the group. The first of these two selective factors has been described by several writers (cf., e.g., 40, 78). That the second is also likely to operate, especially in studies in which less intensive criteria of classification are employed, is suggested by an inspection of the data on average IQ differences reproduced above. It will be noted that only 39 pairs of like-sex fraternals were identified, in contrast to 84 pairs of unlike-sex fraternals. Although in general the number of like-sex and unlike-sex pairs should be roughly the same, in this study less than half as many like-sex as unlike-sex fraternals are listed. None of the unlike-sex pairs could be mistaken for identicals, whereas such a confusion could occur with those like-sex fraternal pairs who were closely similar.

Relatively few surveys of twin resemblance in *special aptitudes* have been conducted. What data are available suggest that in these characteristics, too, identical twins are much more alike than fraternals. In both types of twins, however, the resemblance in special aptitudes is much less than in tests of general intelligence. On a series of tests of motor skills given to 46 pairs of fraternal and 47 pairs of identical twins, the correlations averaged .43 for fraternals and .79 for identicals (39). On the Minnesota Spatial Relations Test, a paper-and-pencil group test of the ability to visualize spatial relations, a correlation of .28 was found within 33 pairs of fraternal twins, and .69 within 29 pairs of identical twins (4).

In *personality tests*, twin correlations tend to be lower than in tests of ability. Moreover, in the personality area, twin correlations are more nearly alike for fraternal and identical twins than they are in

the case of intellectual functions. The degree of twin resemblance in personality characteristics also varies widely with the specific aspect of personality under consideration. All these findings are in line with the results reported in the preceding chapter on parental and sibling correlations in personality tests.

On the Bernreuter Neurotic Inventory, correlations of .63 for identical and .32 for fraternal twins were obtained (8). Another test in the same general area, the Woodworth-Mathews Test of Emotional Instability, gave an identical twin correlation of .54 and a fraternal twin correlation of .28 (26). On tests of other personality characteristics, such as dominance or self-sufficiency, the test correlations tend to run lower (8). The Strong Vocational Interest Test yielded correlations of only .50 for identical twins and .28 for fraternals (8). Although from time to time selected cases of very close resemblance in the personalities of twins are reported, equally striking cases of differences can be found. For example, in a study of ten pairs of fraternal and two pairs of identical twins located within a college population, tests indicated *less* agreement between twins than between siblings in such characteristics as self-sufficiency, introversion-extroversion, social adjustment, and masculinity-femininity (49). Some evidence was found in the same study that a pair of twins may tend, somewhat more often than siblings, to develop *opposite* trends in dominance and submission. The implications of such findings will be examined in the following section, in connection with the social interaction of twins.

Mention may also be made of the various reports of similarity in *crime* and in *insanity* [4] among twins, a topic which has proved especially alluring to popular writers. In one survey covering 13 criminals who were known to have an identical twin, 10 cases were found in which the other member of the pair also had a criminal record. Out of 17 fraternal pairs included in the same survey, only 2 showed both twins to have been convicted of crimes (31, p. 46). Attempts have likewise been made, in tracing the careers of identical twins, to find an "equivalence" between certain forms of illegal and certain forms of legal behavior (31, 67).[5] The behavior of the twins was considered equivalent if it appeared to stem from a common "inher-

[4] Data on the incidence of *specific psychoses* among twins, as well as among siblings and parents and children, will be presented in the special chapter on abnormality, Chapter 16.

[5] Cf. also 78 for other sources.

ent tendency," though differently expressed. Obviously such interpretations leave the way open for much subjective bias. It should also be noted that common environmental factors, likely to be greater for identical than for fraternal twins, could account for much of the observed similarity of behavior. Even the knowledge that one is the identical twin of a criminal might play an important part in determining the individual's own attitude as well as the reactions of others toward him. Finally, as frequently happens in studies of isolated cases, other instances can be found to illustrate the opposite conclusion. Cases are on record in which one member of a pair of identical twins was either criminal or clearly psychotic, while the other gave every indication of remaining normal (cf., e.g., 24).

THE ENVIRONMENT OF TWINS

Fraternal *versus* Identical Twins. All investigators agree in finding identical twins more nearly alike than fraternal twins in abilities, as well as in most other behavior characteristics which have been studied. Identical twins have identical heredity; fraternal twins do not. Can we, then, conclude that the greater resemblance of the former is the result of heredity? It is not so simple as that. The identical twins' closer similarity of heredity is paralleled by a closer similarity of environment. This fact has received increasing recognition in recent research on twins. On the basis of extensive field study of twins, Carter (8) argues against the assumption that nurture influences are even approximately the same for identical as for fraternal twins. He writes:

> Such an assumption seems untenable to anyone who has had much contact with twins in their own social environment, for it is quite evident that the environments of identical twins are on the average more similar than those of fraternal twins. The identical twins obviously like each other better; they obviously have the same friends more often; they obviously spend more time together; and they are obviously treated by their friends, parents, teachers, and acquaintances as if they were more alike than fraternal twins are (8, p. 246).

Many other investigators lend support to such a conclusion. It is clear that fraternal twins are often quite unlike in body build, general health, eye and hair color, muscular strength, and many other physical characteristics (70). One twin may be ugly and the other handsome;

one sickly and the other hale and vigorous. The effect which these physical differences will in turn have upon the twins' relations to their environment may be very far-reaching (25, 78). Each twin will, by virtue of his physical characteristics, automatically "select" different features from the same environment. Actual observation has repeatedly shown that the amount of shared experience of fraternal twins is less than that of identical twins. For example, in a questionnaire (75) answered by 70 pairs of identical twins, 69 pairs of like-sex fraternals, and 55 pairs of unlike-sex fraternals, 43% of the identicals reported that they had never been separated for more than one day. Among the like-sex fraternals, only 26% reported this to be true. Identical twins more often share the same room at home, have the same chum, and are treated more similarly by their families and associates (30). In fact, it is not uncommon for one twin to be mistaken for the other, especially in childhood. All this furnishes an interesting illustration of the indirect influence which physical similarities may exert upon behavior. These similarities, which are themselves largely determined by hereditary factors, may in turn alter the individual's environment in such a way as to affect his behavior development.

A word may be added in this connection regarding comparisons between like-sex and unlike-sex fraternals, as well as between fraternal twins and siblings. The greater similarity in test performance generally found for like-sex than for unlike-sex fraternals could result from either hereditary or environmental factors. On the side of heredity, it will be recalled (Ch. 4) that the presence of sex-linked, sex-influenced, and sex-limited factors may introduce a number of hereditary differences between unlike-sex children of the same parents, which are not present in like-sex children. On the side of environment, it is apparent that the effective environments of a boy and a girl are more dissimilar than would be the case for two boys or two girls. Thus the differences in the results obtained with like-sex and unlike-sex fraternals do not lend themselves to unambiguous interpretation. Any differences in degree of resemblance between fraternal twins as a group and siblings as a group, however, can logically be attributed to the greater environmental similarities of the twins.[6] On the basis of heredity, fraternal twins should be no more alike than ordinary siblings. But their environments will tend to be more similar.

[6] Except in so far as the selective factors discussed in the preceding section may have operated.

This is obviously true of prenatal and natal conditions. Moreover, being of the same age, the twins will be exposed to any changing influences in the home environment at the same stage of their development. The attitudes of parents and associates toward the children, as well as the attitude of the children toward each other, are also likely to differ in the two situations.

Prenatal and Natal Factors. It is obvious that any differences noted within a pair of identical twins must be the result of environmental factors. When identical twins reared in the same home show conspicuous dissimilarities in development, the possible role of prenatal factors or of birth injuries is suggested. That prenatal conditions may produce deficiencies in one twin while the other develops normally is quite consistent with what is known regarding the embryology of twinning. During prenatal life, the twins are competitors for the available supply of nourishment. Sometimes one twin loses out completely and fails to survive, while the other develops at his expense. When the inequality is milder, both are born, but one may be weaker than the other.

An example of the possible operation of prenatal and natal factors in producing differences between identical twins is to be found in the occurrence of feeblemindedness. In a survey of several feebleminded institutions, Rosanoff *et al.* (51) located 126 persons known to have an identical twin. In the majority of these cases, the other twin was also feebleminded or showed some other abnormal condition such as epilepsy, birth paralysis, or behavior difficulties. In 11 pairs, however, no defect was found in the other twin. Since the abnormal condition in the defective twin in these pairs appeared early in life, the probability of birth injuries or prenatal factors is strongly suggested. Some investigators (cf., e.g., 51) consider cerebral birth injuries to be a relatively common, unsuspected cause of mental deficiency. An injury too mild to attract notice at the time may nevertheless be sufficient to interfere with normal intellectual development later on. This point of view has been most vigorously championed by Rosanoff (51). Since twins tend on the whole to be born prematurely—when they are relatively small and weak—they are especially subject to birth injuries. Rosanoff (51) estimates that conditions favoring birth injuries are about eight times as frequent in the birth records of feebleminded persons as in the general population.

The prenatal and natal conditions surrounding the development

of twins have also been cited in explanation of the finding that twins as a whole tend to be *intellectually inferior to single-born children* (8). Such retardation has been noted from the preschool level (cf., e.g., 10) through high school. In a survey of 412 pairs of twins located in a high school sample of 119,850 students, the average percentile score on the Henmon Nelson Test of Mental Ability was 39.73 for the twins and 50 for the rest of the group[7] (7). Further corroboration is furnished by the results of intensive follow-up studies of triplets and quadruplets, as well as by the detailed observations of the widely publicized Dionne quintuplets (2, 3, 8, 16). It can be argued that the larger the number of individuals competing for survival in the uterine environment, the more severe the handicaps imposed upon all of them. The observed facts regarding the intellectual development of multiple-birth children appear to lend some support to such an hypothesis. The fact that multiple-birth children are often born prematurely could also account in part for their retardation, since they are actually at an earlier stage of development than their age indicates. It is doubtful, however, whether this factor has a significant effect upon intellectual development in later childhood; its influence is probably limited in large part to early sensory and motor development.

The Language Development of Twins. A note of caution must be sounded against the too facile acceptance of these structural explanations of intellectual retardation among twins. It should be noted that the observed retardation is generally most marked in the acquisition of language. This in turn has an important bearing upon other forms of subsequent intellectual development. The backwardness in language may result at least in part from the presence of two (or more) children of identical age in the same family. It is a common observation that twins frequently form a relatively self-sufficient unit, and consequently have less need for contact and communication with other children and adults. It is just these contacts, however, which provide powerful incentives and opportunities for learning to talk. That twins may be physically retarded or handicapped because of the physiological conditions of twin development seems quite apparent. But that the same conditions provide a sufficient explanation of their intellectual retardation is not conclusively established.

[7] Difference/$PE_{diff.}$ = 10.37; the statistical significance of this difference is therefore very high.

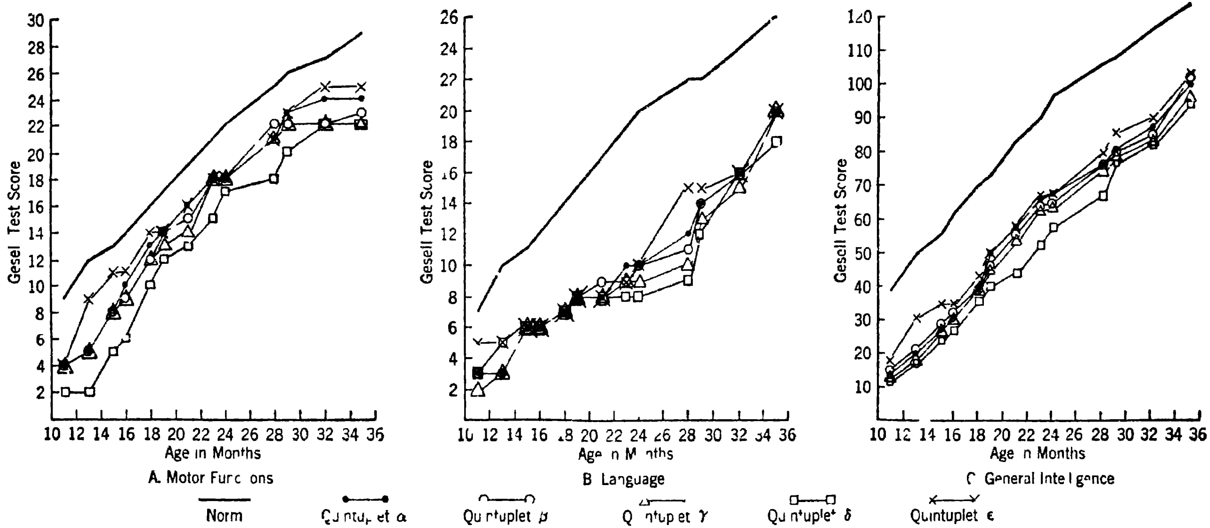


Fig. 65. Development of the Dionne Quintuplets in Motor Functions, Language, and General Intelligence (Data from Blatz *et al.*, 3, Graphs 1, 2, 3)

Let us examine some of the specific data on language development among twins. Figure 65 shows the development of the Dionne quintuplets in (a) total intelligence test scores, (b) motor functions, and (c) the acquisition of language. It is apparent that the greatest retardation is in language and the least in motor functions. The indices of "general mental development" occupy an intermediate position, probably because of their composite nature. Although the quintuplets were born two months before the normal term, and although the possibility of fetal handicaps exists, it is unlikely that such conditions would produce a more marked retardation in language than in motor development. In his discussion of the linguistic retardation of the

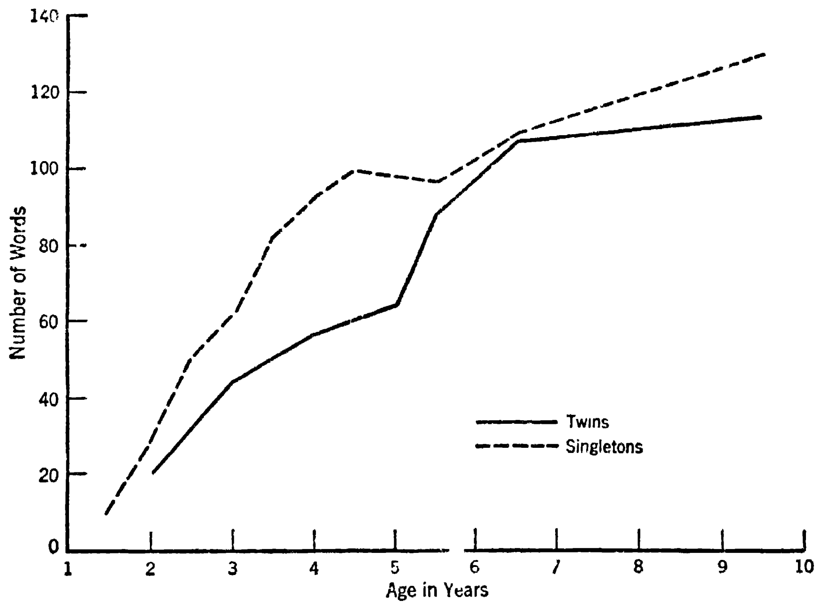


Fig. 66. Linguistic Development of Twins and Singletons. (From Davis, 10, p. 136.)

quintuplets, Blatz (2) calls attention to a number of likely environmental factors. Since most of their wants were anticipated by ever vigilant attendants, the children had little need to communicate with adults. They had little to tell each other, since they shared most experiences. By age three, moreover, they had developed a number

of mutually intelligible gestures and cries to express their feelings among themselves.

Group surveys of triplets (27) and twins (10, 11) have yielded similar results. Special systems of communication, through gestures and vocal cues, are frequently developed by twins out of their common experiences. The need for acquiring the language of adults is thus reduced. Specific indices of language development, such as length of response or number of different words used during a standard observation period, show consistently more retardation than is found in the total IQ of such children. The extent of this linguistic retardation is illustrated in Figure 66. Bringing together the data of several investigators, this figure shows the average number of different words used during the examination period by twins and singletons (singly born children) between the ages of 1½ and 9½. The difference between the two groups appears to be largest from ages 3 to 5, and decreases somewhat during school ages.[8]

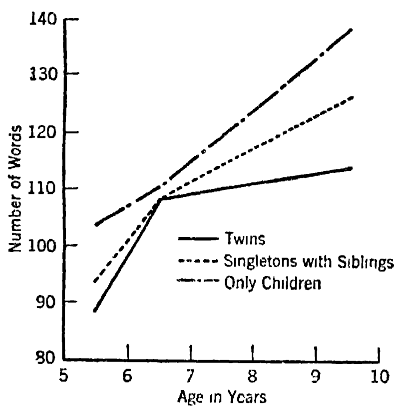


Fig. 67. Comparison of the Language Development of Twins, Singletons with Siblings, and Only Children. (From Davis, 10, p. 112).

That the contact of twins with each other is a major factor in their linguistic retardation is further suggested by the finding that *only-children* are definitely superior to children-with-siblings in every phase of linguistic skill (10). In fact, singletons-with-siblings resemble twins in many phases of their language development more closely than they resemble only-children. In Figure 67 will be found the number of different words used during a test period by twins, singletons-with-siblings, and only-children at ages 5½, 6½, and 9½. It will be noted that the singletons-with-siblings are somewhat closer to the twins than to the only-children at the first two age levels, and about midway at the third. Also relevant is the finding that

[8] The very small difference at age 6½ may be due to the relatively small number of cases examined at this age.

children who spend more time with adults tend in general to be linguistically superior (38).

In conclusion, the retarding effect of the "twin environment" upon language development seems to be quite clearly demonstrated. Linguistic retardation in turn has far-reaching implications for all intellectual development. Not only is language necessary as a means of communication in most human learning, but linguistic symbols themselves play an important part in problem solving and in the more abstract and complex human intellectual functions.

Social Interaction. The social reactions of twins toward each other provide a promising field of investigation in themselves. Many observers have called attention to the *specialization of "roles"* which twins often seem to work out by a tacit mutual agreement (47, 78). Such a division of labor—observed especially among identical twins—makes for more harmonious relationships and economy of effort. Thus one twin may be the spokesman for the pair in encounters with other persons, showing more interest in people and responding more actively to them. Frequently one twin is the dominant member of the pair, tending more often to lead and to make the decisions for both (11, 78). Such a differentiation of roles may originally arise from slight differences in size and strength, which may have been prenatally established. The parents' efforts to discover and emphasize any distinguishing mark between the twins may be a further source of differentiation. In some cases, minor chance happenings may initiate the difference, which is then willingly accepted and developed by the twins as a matter of convenience.

Such a divison of roles, continued and augmented over the years, could account for some of the differences in interests, attitudes, emotional reactions, and abilities sometimes found between twins reared in the same home. For example, the relatively large differences between twins in such traits as dominance, self-sufficiency, introversion, and the like, found in the previously cited study on a college population, are understandable on this basis (49). These findings are supported by more detailed case studies of individual twins (cf., e.g., 42) and are borne out in very interesting ways by the observations of larger multiple-birth groups. The Dionne quintuplets, for example, although reared under as nearly uniform and controlled conditions as any group of children, nevertheless show clear-cut personality and ability differences (2, 3). Yet the conclusive demonstration of their

identical heredity precludes any explanation of such differences in terms of heredity. In commenting upon these findings, Blatz writes:

> It is in the environment, apparently the same for all, that there nevertheless exist subtle yet important differences in the influences bearing on these children—differences of which the social interaction of the five, one upon the other, is the most emphatic yet the most difficult to identify and measure (2, p. 174).

Another vivid demonstration of environmentally determined differences among identical twins is furnished by the Morlok quadruplets [9] of Lansing, Michigan (16). On the Stanford-Binet Intelligence Scale, one of these quadruplets received an IQ of 110, another 101, while the other two occupied intermediate positions. This relationship was consistently maintained on other tests of intelligence or scholastic ability. On the Stanford Achievement Test, for example, the "bright" twin earned a total score of 124 and the "dull" one 96. Each of the sub-tests of the Stanford Achievement Test showed the same relationship. Physical and personality differences among the quadruplets closely paralleled these intellectual differences. The dullest twin was also the smallest and had had a poorer health history throughout childhood. The investigators suggest the possibility of differences in fetal blood supply as a basis for both the physical and the psychological dissimilarities. It may well be that the initial, prenatally determined, physical differences led to a subsequent social diversification of roles among the four sisters, which in turn affected their subsequent emotional and intellectual development. The personality differences among the four are reported to be especially conspicuous. The children have been characterized by their parents as "the boss," "the clown," "the artist," and "the baby," and the investigators report that an outside observer could readily identify the child fitting each of these labels, even from a brief observation.

TWINS REARED APART

Of considerable interest are the case studies of identical twins who were separated at an early age because of death of parents or other

[9] These are the only known living one-egg quadruplets. A series of studies of other quadruplets have been reported by I. C. Gardner and H. H. Newman in the *Journal of Heredity*, 1940, 31, 307-314, 419-424; 1942, 33, 311-314, 345-350; 1943, 34, 27-32.

misfortunes, and were reared in separate homes. About twenty-five such pairs have been located and carefully studied. The most extensive collection of cases has been assembled in an investigation at the University of Chicago, conducted by Newman, Freeman, and Holzinger, a geneticist, psychologist, and statistician, respectively. The principal study (47) covered 19 pairs of identical twins, most of whom had been separated since their first year of life. The actual age of separation in individual cases ranged from two weeks to six years. All the twins had lived apart up to the time of their examination, although in one or two cases they had corresponded or occasionally visited each other. The ages at the time of testing ranged from 11 to 59 years.

Each case was intensively studied through physical measures, psychological and educational tests, and personal interviews. Data were also obtained regarding the foster home and foster parents, educational and vocational history of the twins, health and disease records, and other relevant factors in the subjects' experiential background. A case history illustrating the effects of two fairly dissimilar homes upon a pair of identical twins is summarized below.

Case #4. Mabel and Mary, 29-year-old twins, had been separated at the age of five months and reared by relatives. Mabel had led the life of an active farm woman on a prosperous farm. Mary had lived largely a sedentary life in a small town, clerking in a store during the day and teaching music at night. Mabel had only an elementary school education in a rural school, while Mary had had a complete high school course in an excellent city school. At the time of examination, a vast difference was noted between the twins in intellectual, emotional, and physical traits. Physically, Mabel is described as robust, muscular, and in perfect health, while Mary was underweight, soft-muscled, and in poor general condition; Mabel weighed 138½ lbs., Mary only 110¾ lbs. Intellectually, an equally striking difference was found, but in favor of Mary, whose Stanford-Binet IQ was 106 as compared with 89 for her sister. Even larger differences were obtained in some of the other tests. In personality characteristics, the twins exhibited consistent differences, as determined both by tests and by direct observations. The rural twin tended to be more stolid and stable in emotional responses, to give fewer neurotic reactions, worry about fewer things, and respond less emotionally to stimuli than did the urban-bred twin. Both the physical differences noted above and the contrast between their psychological environments probably account for these personality differences between the twins (47, pp. 187–195).

TABLE 15 *A Comparison of Identical Twins Reared Apart with Fraternal Twins and with Identical Twins Reared Together*

(Adapted from Newman, Freeman, and Holzinger, 47, pp. 72, 97, 344, 347, and Woodworth, 78, p. 19)

Measure	*Mean Difference between Twins* Fraternal (50 pairs)	Identical Together (50 pairs)	Identical Separated (19 pairs)
Height in cm	4.4	1.7	1.8
Weight in lb.	10.0	4.1	9.9
Binet IQ	9.9	5.9	8.2

Measure	*Correlations* between Twins* Fraternal (50 pairs)	Identical Together (50 pairs)	Identical Separated (19 pairs)
Height in cm.	.64	.93	.97
Weight in lb.	.63	.92	.89
Binet IQ	.63	.88	.77

* Certain necessary statistical corrections, for age and for inequalities in IQ range, have been applied to the original correlations given by Newman, Freeman, and Holzinger (47). For further details on these corrections, see 26, 40, and 78.

To obtain comparative data, Newman, Freeman, and Holzinger tested 50 pairs of *identical twins living together* and 50 pairs of *fraternal twins*, also living together and with their own families. Mean differences as well as correlations in height, weight, and IQ for all three groups are shown in Table 15. It will be noted that the average IQ difference between the separated identicals is 8.2, slightly less than the mean difference between fraternals, but slightly larger than that between non-separated identicals. Essentially the same relationship is brought out by the correlation coefficients, the separated identicals falling between the non-separated identicals and the fraternals in closeness of resemblance.

Several writers have tried to draw inferences regarding heredity and environment from a comparison of the degree of resemblance in *physical* and *intellectual* traits in these three groups. In Table 15, the results for height, weight, and IQ are not startlingly different. To be sure, the height and weight correlations for the two groups of identical twins are higher and more nearly alike than are the cor-

responding intelligence test correlations. But we must bear in mind that measures of height and weight are more reliable than measures of intelligence. Chance errors of measurement in intelligence tests would introduce more fluctuations in score, even in retests of the same individual. Such chance fluctuations tend to lower the correlations of IQ's and render any differences between such correlations less significant. In other words, we cannot conclude with certainty from the data in Table 15 that the effect of separation was any greater on IQ than it was on height and weight. At the same time, it should be remembered that such comparisons between familial resemblances in physical and in psychological characteristics, although of some interest in themselves, really tell us little about the problem of heredity and environment. Similar correlations *can* result from different factors, and their similarity is therefore no indication that the same influences have operated.

A more crucial approach is provided by an analysis based upon the *extent of environmental differences* between the two twins in each separated pair. The mere separation of the twins need not in itself lead to differences. It is conceivable, as a matter of fact, that a particular pair of twins reared apart may be more alike than if they had been reared together. If their respective environments are closely similar—although geographically remote—identical twins should respond with considerable uniformity. Their physical likenesses, based upon a common heredity, would in such a case insure like responses to like stimulation. Brought up together, on the other hand, the same two twins might show divergent courses of development owing to the specialization of "roles" discussed in the preceding section. Psychologically, environment is not geography.

In Table 16 will be found individual data on each pair of separated identical twins. The original 19 cases of the Chicago study have been augmented with a twentieth case subsequently reported by Gardner and Newman (15). The IQ differences in the last column indicate the excess in favor of whichever twin received the better education. An examination of these IQ differences suggests that, on the whole, they are not random differences such as might result from fortuitous factors, but rather tend to favor the better educated twin quite consistently. If we restrict our comparisons to the five pairs which present large differences in amount of schooling (first five cases in Table 16), the mean IQ difference in favor of the better educated twin is 16

TABLE 16 *Individual Data on Twins Reared Apart*

(Adapted from Woodworth, 78, p. 23, with additional data from Newman, Freeman, and Holzinger, 47)

Case Number	*Sex*	*Age at Separation*	*Age at Testing*	*Environmental Differences*: 1. In Years of Schooling	2. In Estimated Educational Advantages	3. In Estimated Social Advantages	4. In Estimated Physical Advantages	*Twin Difference in IQ*
11	F	18 mo.	35	14	37	25	22	24
2	F	18 mo.	27	10	32	14	9	12
18	M	1 yr.	27	4	28	31	11	19
4	F	5 mo.	29	4	22	15	23	17
12	F	18 mo.	29	5	19	13	36	7
1	F	18 mo.	19	1	15	27	19	12
17	M	2 yr.	14	0	15	15	15	10
8	F	3 mo.	15	1	14	32	13	15
3	M	2 mo.	23	1	12	15	12	−2
14	F	6 mo.	39	0	12	15	9	−1
5	F	14 mo.	38	1	11	26	23	4
13	M	1 mo.	19	0	11	13	9	1
10	F	1 yr.	12	1	10	15	16	5
15	M	1 yr.	26	2	9	7	8	1
7	M	1 mo	13	0	9	27	9	−1
19	F	6 yr.	41	0	9	14	22	−9
16	F	2 yr.	11	0	8	12	14	2
6	F	3 yr.	59	0	7	10	22	8
9	M	1 mo	19	0	7	14	10	6
20*	F	1 mo.	19	0	2	[illegible]	[illegible]	−3

* The first 19 cases are from Newman, Freeman, and Holzinger (47); Case 20 was later added to the collection by Gardner and Newman (15). Five additional cases, studied by other investigators, are cited in the text. They have not been included in the table since the data do not lend themselves to comparable evaluations in terms of the above categories.

points. It will be noted that in the remaining cases the differences in schooling are small or non-existent. In so far as schooling may affect the IQ, then, these remaining twins would not be expected to differ much. And the differences are, in fact, small.[10] If the cases with similar educational opportunities, where little or no difference is expected, are averaged with those showing clear-cut educational differences, then the possible effects of this environmental factor are diluted and underestimated. A composite figure based upon all these cases, in which specific conditions varied so widely, would only obfuscate the results.

On the basis of the case material, the environments of the separated twins in the Chicago study were rated by five judges for degree of intra-pair difference in educational, social, and physical or "health" advantages, respectively. These ratings are also given in Table 16. Since each of the judges used a 10-point scale, their combined ratings for each characteristic could have a maximum value of 50. The higher the rating, the greater the estimated difference in environmental advantages between the two twins in each pair. These ratings show interesting correspondences with the observed differences in intellectual, emotional, and physical characteristics of the twins. Thus a correlation of .79 was found between the discrepancies in educational advantages and the discrepancies in IQ within each pair of twins. IQ differences correlated .51 and .30, respectively, with judged differences in "social" and in "physical" environments.[11] Twin discrepancies in body weight, on the other hand, correlated .60 with discrepancies in the "physical" environments.

It should be noted, moreover, that the types of homes into which the twins in any one pair were placed rarely differed very much. If an experiment were being designed to test how far environment may affect, for example, the IQ, the twins would obviously be placed in as different homes as possible. But in the actual placement procedures followed, the reverse tendency probably operates. The placement of children in foster homes tends to be selective, an effort being made to place the children with families similar to their own. In a number of twin pairs, the children were adopted by relatives. This

[10] Two cases, 1 and 8, show evidence in their histories of a possible prenatal handicap, which tended to make the twins unlike. If these two cases are omitted, the remaining cases show an insignificant mean difference of less than 2 IQ points.

[11] The correlation of .30 is not statistically significant; the other reported correlations meet the usual standard of significance.

would certainly make for greater similarity in socio-economic, educational, and other characteristics of the two foster homes than would be the case between two families picked at random.

Five additional pairs of separated identical twins, studied by other investigators, tend to corroborate the major findings of the Chicago survey. A case reported by Muller (44) in 1925, the earliest on record, showed a negligible difference between the twins on two group tests of intelligence. Although one twin had had only 4 years of formal schooling and the other 13, the educational levels of the foster parents and the socio-economic levels of the two homes were closely similar. Both girls are said to have read "voraciously," a fact which may have helped to counteract the differences in schooling.

More recently, two pairs of separated identical twins have been studied in this country (6, 68), and two pairs in Great Britain (52, 79). In none of these four pairs was there any difference in amount of schooling received. Other differences in opportunities for intellectual development, as suggested by the descriptions of the homes, foster parents, or type of education, appear to have been either minor or counterbalanced within the pair. For example, if one member of the pair was handicapped by frequent changes of schooling, she had the advantage of higher socio-economic level of the foster home, in comparison with her twin. In three of these four cases (6, 52, 79), differences in intelligence test scores were uniformly small and insignificant, although a number of differences in attitudes, social conformity, and other personality characteristics were noted.

An interesting divergence in special aptitudes is presented by the fourth recently reported case, that of a pair of British twins separated at 3 months and reared apart until the age of 16 years (79). Although both had received the same amount of schooling, one twin had a Stanford-Binet IQ of 125, the other of 106. The twin with the lower Stanford-Binet IQ, however, excelled consistently in performance tests and in tests of mechanical aptitude; these differences amounted to as much as two years in mental age and nearly 30 percentile points, respectively, in the two types of tests. Moreover, the twin with the higher Stanford-Binet IQ (and lower mechanical and performance test scores) was inferior to his co-twin in height, weight, and general health. The possible effect of prenatal and postnatal environmental factors upon physique and health, which in turn might

influence the divergent development of interests and aptitudes, is suggested by these results.

To the case studies of identical twins reared apart may be added the educational experiment reported by Schmidt (53). In this investigation—the major portion of which was discussed in Chapter 8—9 pairs of identical twins were included within the total group. It will be recalled that in this study a number of children originally classified as feebleminded were able to make a relatively normal educational, intellectual, and vocational adjustment, as a result of a specially designed three-year educational program. Among the 9 pairs of identical twins, one member of each pair participated in the special program, while the other remained either in regular public school classes or in the usual ungraded classes provided for backward children in the public education system.

The average IQ of the twins in the experimental group rose from 54 to 92 in the course of the special program; the control twins showed a negligible change from 61 to 59 during the same period. It will be noted that, initially, the average IQ of the control twins is higher. This initial difference results from the fact that when the two twins in any pair differed appreciably in IQ, the one with the *lower* IQ was chosen for the experimental program. Because of this procedure, it is unlikely that the final advantage in IQ in favor of the experimentally trained twins could result from possible prenatally determined structural advantages. On the other hand, the regression effect (cf. Ch. 8) would account for a slight tendency for the initially higher control group to lose and for the initially lower experimental group to gain. Regression alone could hardly account for a large part of the observed rise in IQ in the experimental group, however, especially in view of the high reliability of the Stanford-Binet during the elementary school ages covered by the study. Emotional readjustment may have played a significant part in the improvement in intellectual, educational, and social performance of the experimental subjects, since such readjustments were an integral part of the program. Individual remedial instruction, better work habits, and improved attitudes and motivation all undoubtedly contributed to the gains in IQ in individual cases.

Methodologically, the twin analysis in Schmidt's investigation falls midway between the case studies of twins reared apart and the training experiments by the method of co-twin control cited in Chapter 6.

It is broader in scope, of longer duration, and concerned with more complex behavior functions than the previously cited co-twin-control studies. At the same time, the separation was not so complete nor did it begin so early as in the case studies reported in the present chapter. During the three-year period of the experimental training, however, the twins in the Schmidt study were probably exposed to more dissimilar stimulation than was true of most of the accidentally separated pairs studied by others. There is undoubtedly need for more experimentally controlled co-twin studies of the development of *complex behavior functions*. Such an approach should prove to be the most promising of all those employed in the study of twins.

FOSTER CHILDREN

How Well Do Foster Children "Turn Out"? The development of children reared in foster homes is of considerable interest for practical as well as theoretical reasons. Is there any basis for the popular belief that adopted children "turn out badly"? On the whole the answer appears to be "No," although the contributing factors are too many and too complex to permit a categorical denial. Follow-ups of a group of 910 adopted children indicated that as adults the majority had made a satisfactory vocational and social adjustment (73, 78). A little less than a fourth were judged unsatisfactory in their adjustment because of educational backwardness, shiftlessness and dependency, or delinquency and crime. This proportion is larger than that in the general population, but smaller than would have been expected if the children had been reared in the unfavorable environments from which they were frequently taken. Within the adopted group, a relationship was found between the quality of care and child training provided by the foster home and the number of foster children judged to have made a successful adult adjustment. In the homes rated "excellent" in this regard, 87% of the foster children fell in the satisfactory adult category; in the homes rated "poor," only 66% were so classified.

On intelligence tests, foster children as a group tend to fall somewhat below "own children" brought up in comparable homes (5, 22, 78), but above the average of the general population. At least two factors may account for the latter difference in favor of the foster

group. First, placement agencies as well as foster parents tend to choose the most promising children for placement in foster homes, while the more poorly qualified tend to remain under orphanage or boarding-home care. Secondly, the same type of selection occurs with reference to foster homes, the more undesirable homes at the lower end of the distribution being disqualified for adoption purposes. Foster children as a group are thus reared in homes superior to the general average. It is also likely that foster parents, on the whole, have a relatively strong interest in children; otherwise they would not have gone out of their way to adopt one.

Why are foster children less successful—in intelligence test performance as well as in adult achievement—than other children reared in the same type of home? A number of psychologists put the burden of explanation upon unknown "hereditary influences." Presumably this means *genetically determined structural limitations* on behavior development. Such limitations may play a significant part in individual cases, but little or no direct information is available regarding what they are. Part of the explanation, on the other hand, may be provided by *prenatal and natal environmental factors*. Such conditions as diet and medical care of the mother during pregnancy and parturition are probably inferior, in general, for the foster group. That these conditions may affect the structural development—and indirectly the subsequent behavioral development—of the child is being increasingly recognized (62, 71). On the basis of extensive observations at the Fels Research Institute, for example, Sontag writes:

> Contrary to earlier opinion—the progress of a fetus and of an infant is considerably influenced by the quality of the diet of his mother during the gestation period. . . . there is increasing evidence of the tremendous importance of maternal nutrition and variations in endocrine function in determining the physique, physiology, and progress of the neonate. It is, it seems to me, self-evident that the physical and physiological adequacy of the neonate are in turn critical factors in his emotional and social adaptation during infancy and therefore throughout life (62, pp. 151–154).

When the child has lived with his own family or in an institution for several years prior to adoption, the possible influence of such *early home environment* needs, of course, to be taken into account. A further factor to consider is the *nature of the family relationship in foster*

homes. The attitude of foster parents toward a child may differ in some essential ways from that of own parents. In some cases, the contact of foster parent and child may not be so close or intimate as that of a child and his natural parents. The child himself, when he knows of his adoption, may react differently toward his foster parents than he would toward his own parents. Social expectancy may also complicate the situation. Parents as a rule expect their own children to resemble them in intellectual and emotional development, and this expectation may be manifested in their behavior toward the child, as well as in the attitudes of other relatives and associates. As the child develops, his observers repeatedly call attention to points of family resemblance, real or imagined; he is frequently reminded of ancestral characteristics, which are held up to him as his heritage. Social influences of this sort are absent or greatly minimized in the case of foster children. It would be difficult to estimate what subtle motivational differences may arise as a result of such differences in social expectancy, and what effect the motivational factors may in turn have upon the subsequent course of intellectual development of the child.

Foster Children and the "Nature-Nurture" Question. To psychologists, foster children have provided one more approach to a possible determination of the contributions of heredity and environment to intellectual development, and a number of investigations have been especially designed with this problem in mind. Three major types of analysis have been employed for this purpose: (1) comparison of foster family and own family resemblances; (2) study of the relationship between foster child's IQ and level or quality of the foster home; and (3) determination of change in child's IQ following adoption and residence in the foster home. The four most extensive investigations are those conducted by F. N. Freeman and his associates at the University of Chicago (13), Burks at Stanford (5), Leahy at Minnesota (35), and Skodak and Skeels at Iowa (59, 60, 61). Two of these studies emphasize the contribution of heredity and the other two the contribution of environment. A brief examination of their procedures and findings will show that their discrepancies are more apparent than real.

Burks (5) administered the Stanford-Binet to 214 foster children and their foster parents, as well as to 105 control children and their own parents. The control group was closely equated with the foster

group in age of children and parents, educational and occupational level of the parents, and cultural characteristics of the home. All subjects were white, English-speaking, and quite homogeneous in national and cultural background. Each foster child had been legally adopted by a married couple, the two foster parents being alive and living together at the time of the study. Each control child was likewise in a home in which both parents were living. Only foster children who had been placed in the foster home under the age of 12 months were included, the average age of placement being three months. At the time of testing, the children ranged in age between 5 and 14.

The correlations between Stanford-Binet mental ages of the parents and IQ's of the children, for both foster and control groups, are presented below. The correlations between child's IQ and a composite cultural index of the home are also given.

Correlation between Child's IQ and:	*Foster*	*Control*
Father's MA	.07	.45
Mother's MA	.19	.46
Cultural index of home	.25	.44

Since the resemblances in the control group, attributable to heredity plus environment, are consistently closer than those in the foster group, attributable to environment alone, Burks concludes that heredity is much more important than environment in determining individual differences in intelligence. She estimates that

> The maximal contribution of the best home environment to intelligence is apparently about 20 IQ points, or less, and almost surely lies between 10 and 30 points. Conversely, the least cultured, least stimulating kind of American home environment may depress the IQ as much as 20 IQ points (5, p. 309).

In the investigation by Leahy (35), the same general procedure was followed as in Burks' study, with certain improvements. The Otis Self-Administering Test (Intermediate Form) was substituted for the Stanford-Binet in testing the parents, this test being better adapted to the adult level than was the Stanford-Binet. The matching of the experimental (adopted) and control groups was done very meticulously, each adopted child being paired with a control child of the

same sex, age (within six months), father's occupational level, and father's and mother's schooling.[12] The age of adoption was even lower than in Burks' study, all children having been placed in the foster homes at six months of age or younger.

All other conditions in both experimental and control groups were the same as in Burks' study, except that Leahy's foster group included only illegitimate children, while less than 80% of Burks' group were illegitimate. This difference probably introduces certain selective factors. Illegitimate children come from families of varied socio-economic and intellectual background; whereas other foster children, adopted because of incompetence or poverty of parents, or similar reasons, tend obviously to come from lower-level families. On this basis alone, illegitimate children as a group would be expected to be average in heredity. But it has also been shown (34) that the parents of those illegitimate children who are placed for adoption are of higher average educational and vocational level than those parents who retain their illegitimate children. The difference probably results from the greater sensitivity to social disapproval among persons in the higher educational and socio-economic levels, who would thus be more reluctant to retain an illegitimate child. This additional selective factor might make the group of *adopted* illegitimate children superior in heredity to the general population. It has also been found that those illegitimate children adopted at *earlier ages* tend to come from the highest socio-economic levels (34).

All these factors suggest that a group of illegitimate children adopted before the age of six months, such as that studied by Leahy, ought to be superior from the standpoint of heredity. Leahy's group did in fact excel in IQ, averaging 110.5. This average is obviously higher than that of the general population, and it is also higher than the IQ of other groups of foster children previously tested. The difference could logically be attributed to heredity, on the basis of the selective factors discussed above. But it should also be noted that a certain amount of selective placement occurs, the children whose own parents are better educated and socially superior being routed by the placement agencies into superior foster homes. The intellectual superiority of such a group might thus be the result of their having been placed in better foster homes.

[12] The last three categories apply to own parents in the control group and to foster parents in the experimental group.

The principal comparisons made by Leahy, paralleling those of Burks, are given below.

Correlation between Child's IQ and:	*Foster*	*Control*
Father's Otis score	.19	.51
Mother's Otis score	.24	.51
Cultural index of home	.26	.51

Various other comparisons were made which led the author to conclude, with Burks, that heredity is the major influence in the determination of intellectual level.

In interpreting the correlations reported by Burks and Leahy, a number of factors must be taken into consideration. First, as previously mentioned, intra-family relationships may not be strictly comparable in foster and own homes. The child's knowledge of adoption may affect his attitude toward his foster parents and foster siblings, as well as his self-confidence and his accomplishment. In Burks' group, 35% of the children knew of their adoption, and in Leahy's group 50%. The parents, of course, always know of the adoption, and their reactions toward the child may be affected by such knowledge in countless ways. Of some relevance in this connection is the finding that two unrelated foster children reared in the same home tend to resemble each other more closely in IQ than an adopted and an own child reared together. To be sure, the obtained differences in correlation are slight and the groups of subjects available for such comparisons too small for conclusive results. But it is interesting to note the consistency of this finding in different studies (5, 13, 35).

A second consideration is the role of natal and prenatal factors, which has also been previously discussed. Such factors, although environmental in nature, would tend to increase the resemblance of children to their own parents, as contrasted with their resemblance to foster parents. Mothers who are intellectually and socially or economically inferior would also be more likely to provide inferior prenatal care through ignorance, irresponsibility, or poverty. It might seem that prenatal environmental factors could not account for resemblance to own fathers. More careful consideration, however, will show that the father's educational, vocational, and economic level will also in part determine the quality of medical, dietary, and other conditions affecting the mother.

A final point should be noted. In several respects, Leahy's study is better controlled than Burks' and consequently its results should be regarded as more conclusive. There is, nevertheless, one serious disturbing factor in Leahy's data. There are several indications that, despite the care with which the foster and control groups were matched in parental education and occupation, the cultural levels of the foster and control homes were not truly comparable (74). The average over-all environmental rating was 137.9 for the foster and 118.7 for the control homes; the corresponding SD's were 54.3 and 59.6, respectively. In other words, the foster homes were on the whole superior and *more uniform* than the control homes. Such uniformity would obviously serve to decrease the contribution of home environment to *individual differences* in IQ. In fact, if home environment plays a significant part in intellectual development, we should expect the foster children in this group to be more alike in IQ than the control children. Such was indeed the case, the foster IQ's having an SD of 12.5 and the control an SD of 15.4. Further evidence for the greater cultural homogeneity of the foster homes is furnished by the average environmental ratings of homes in different occupational categories. The foster homes in which the father is a semi-skilled or a day laborer are much less inferior than are the control homes in the corresponding occupational category, the average environmental ratings being 74.7 and 40.1, respectively. Apparently, matching father's occupation was not a sufficient control, since within the same occupational level those homes which are approved for adoption purposes tend to be at the upper end of the distribution. What is even more important is that homes in different occupational categories were more alike in the foster than in the control group. It is not surprising, in view of this situation, to find that the IQ's of the foster children also differed less from one occupational category to the other than did the IQ's of the control children.

The investigation of Freeman *et al.* (13) employed a wider variety of approaches but was less well controlled in certain important respects than the Burks and Leahy studies. One of its principal weaknesses is that the age of adoption was much higher, averaging four years for the entire group of 401 foster children tested. The subjects were somewhat more heterogeneous in national, racial, and socioeconomic background than in the other two studies. The children were tested with the Stanford-Binet and the International Group

Mental Test,[13] the foster parents were given the Otis Self-Administering Test (Higher Form) and a specially constructed vocabulary test covering many fields of knowledge. Field workers collected data on the education, occupation, and cultural level of the foster parents, as well as on the condition of the foster home. Information regarding the natural parents of the foster children was obtained whenever possible through visits, interviews, and examination of case records.

In order to facilitate various comparisons, the children were classified into four overlapping groups, and the data analyzed separately for each group. Group I, the *pretest group,* consisted of 74 children who had been tested before adoption and who had lived in the same foster home until the time of the second examination. The average age of these children at adoption was eight years and their average period of residence in the foster home at the time of the study was four years. After residence in the foster home, the average IQ of this group showed a small but fairly reliable gain from 91.2 to 93.7.[14] When this group of 74 children was divided into those adopted into the better and those adopted into the poorer foster homes, the former showed an average rise of 5 IQ points, while the latter showed no change. Similarly, those children adopted earlier showed more improvement than those adopted later. Although average gains were small, the results of these various comparisons tended to be mutually corroborative.

Group II, the *sibling group,* was composed of 125 pairs of siblings, each adopted into a different foster home and separated for a period of 4 to 13 years. The average age at which the siblings were first separated was 5 years-4 months. In contrast to the correlation of about .50 ordinarily found between siblings reared in the same home, the IQ's of these separated siblings correlated only .25. The scores of 63 siblings adopted into homes which received significantly different cultural ratings correlated only .19; those of siblings adopted into similar foster homes correlated .30. These discrepancies in correlation are all the more impressive when it is recalled that the siblings had lived together during the important years of early childhood.

The third group included all *foster siblings,* i.e., two unrelated children living in the same home. This in turn was subdivided into a group

[13] A non-language and relatively "culture-free" intelligence test (cf. Ch. 21).

[14] The mean gain is slightly larger than 3 times its probable error. Freeman estimates that the net average gain is 7.5 IQ points after allowance is made for inaccuracies in the standardization of the form of the Stanford-Binet in use at the time of the study.

of 40 pairs consisting of a foster child and an own child of the foster parents, and a group of 72 pairs of unrelated foster children. In the former, a correlation of .34 was found between the IQ's of the two children in each pair; in the latter, the correlation was .37. It will be noted that these correlations are higher than those between true siblings adopted into different foster homes.

Finally, all the children were included in one composite group of 401 cases. This composite, labeled the *home group* by Freeman, was employed chiefly in making general comparisons between the foster child's intellectual and social development, on the one hand, and such factors as foster parents' intelligence and cultural level of the foster home, on the other. In the entire group, a correlation of .48 was found between child's IQ and cultural rating of the foster home. This correlation rose to .52 when only children adopted under the age of two years were included. Presumably these children, having lived in the foster home from a younger age, showed the influence of the foster home more clearly in their intellectual development. The correlation of child's IQ with foster father's Otis score was .37 (N = 180), and with that of the foster mother .28 (N = 255). These and other similar correlations suggest the importance of home environment in intellectual development.[15]

The principal difficulties in the way of an unambiguous interpretation of the findings of the Chicago study are the *selective placement* of foster children and the possible *unreliability of the initial IQ's* in the pretest group. It is a well-known policy of placement agencies to "fit the child to the home." Moreover, the more intelligent foster parents may themselves be more concerned with the intellectual level of a child whom they are considering for adoption. While the less intelligent foster parents would hardly demand or choose a less intelligent child deliberately, they may be less concerned with intellectual level and may base their decision primarily upon other considerations.

Freeman and his co-workers looked into the possibilities of selective placement in their study and were inclined to minimize its effect. The adoption records showed, for example, that health, sex, race, and physical appearance were specified by the foster parents in their applications much more often than intellectual level. When the latter was

[15] The specific results cited in the above discussion are all based on the Stanford-Binet and the Otis tests. The International and the vocabulary tests yielded closely similar results in all the cases in which they were employed.

mentioned, it was only to request that the child be "normal," and this request was made as often by the less intelligent as by the more intelligent foster parents. Moreover, in over 80% of the cases no intelligence test scores were available for the children or for their natural parents. Despite these findings, it should be borne in mind that the placement agency could still use other knowledge about the children to estimate their relative intellectual level. Data on the education and occupation of the child's own parents were probably employed in many instances in choosing a "suitable" foster home (cf. 34). In addition to deliberate attempts to place the more "promising" children in the better foster homes, a certain amount of unsuspected selection may have occurred through the factors of illegitimacy and age, as previously discussed. Younger foster children, it will be recalled, tend to come from superior families. At the same time, superior foster families more often request and adopt younger children. Freeman, for example, found a correlation of .27 between the cultural rating of the foster home and the age of adoption of the child. We may well ask, then, what accounts for the higher IQ's of children in the better foster homes. Are they brighter because their parents were more intelligent, or because they have been reared in a superior foster home? Since the two factors cannot be isolated under existing adoption practices, the question cannot be conclusively answered from the data at hand.

With reference to the gain in IQ shown by the children in the pretest group, it has been suggested (20, 54) that the initial score may have been spuriously lowered by emotional stress. If the child is tested while living in an institution or boarding home, or shortly after arrival in the new foster home, he is likely to be in a period of uncertainty or readjustment. The emotional condition of the child at such a time is probably not conducive to his best performance on an intelligence test. The rise in score upon retesting after several years of residence in a single foster home may thus simply reflect the child's better adjustment to the home situation and his greater freedom from anxiety and other unfavorable emotional conditions.

A longitudinal approach to the study of foster child intelligence is represented by the long-range project conducted by Skodak and Skeels at the University of Iowa (55, 59, 60, 61). Stanford-Binet[16] IQ's were determined periodically on an original sample of 306 children placed in foster homes under the age of 6 months and legally

[16] The Kuhlmann-Binet was used at ages 3½ or lower

adopted. The large majority of the children were illegitimate. The average age of placement was slightly under 3 months; at the time of the first examination, the age of the group ranged from 1½ to 6 years and averaged 2 years. Retests were made at average ages of approximately 4, 7, and 13 years, the number of children available for the last retest being 100. The average IQ of this group of 100 on each of the retests is shown below.

Average age	2	4	7	13
Average IQ	117	112	114	107 (1916 Stanford-Binet)
				117 (1937 Stanford-Binet, Form L)

In the 13-year retest both the earlier (1916) and the revised (1937) form of the Stanford-Binet were used, since the older form had been used at the younger ages. For the most reliable and valid measure of the children's intellectual level, however, the results with the revised form should be considered. The earlier form is likely to underestimate the intelligence of older children, owing to certain inaccuracies of standardization.

Initial IQ's of the foster children showed negligible or zero correlations with intellectual, educational, or occupational level of either foster parents or true parents (when information was available on the latter). With increasing age, however, the correlation between child's IQ and true mother's IQ rose to .44 (63 cases). The low correlations with early IQ may have resulted from the unreliability of preschool testing and from the nature of the functions tested at those ages. As the tests became more verbal and abstract and less sensori-motor in nature, they became better measures of scholastic aptitude or intelligence (cf. Ch. 5). The resemblance of these children to their true mothers may logically be the result of hereditarily structurally imposed limitations of development or of prenatal environmental factors. It may also, however, be the result of selective placement, since the children of the better educated parents were, in fact, placed in the better foster homes.[1]

In their interpretation of these findings, Skodak and Skeels place the major emphasis upon the relatively high average IQ of the foster children in the light of the low intellectual status of their true parents. These interpretations have been the center of much controversy (19,

[1] The correlation (c a) between true mother's IQ and occupational status of the foster father was .35 (41).

41). The highest grade reached in school by the true mothers and true fathers (when data were available) does not appear to be significantly below the average for the general population. Since the majority of the children were illegitimate and all were adopted at an early age, we should expect them to be a relatively superior group. On the other hand, the authors point out that the schooling data tend to overestimate the true parents' ability, since many had been doing poorly in school, were old for their grade, etc. Moreover, the IQ's of 80 true mothers, described as representative of the entire group, averaged 93.[18] Occupational, economic, and social status of the true families was quite low, many of the families being on relief. On the whole, the IQ's of the foster children do seem to be higher than might have been expected from what is known of their true family background, but little more can be concluded.

Foster Children of Feebleminded Mothers. Probably the principal source of contention in the Skodak-Skeels study was the authors' statement that the children of feebleminded mothers were indistinguishable from the rest of the group in their subsequent intellectual development in the foster homes. Although only 16 of the true mothers who were tested fell clearly within the feebleminded range, a certain measure of corroboration is to be found in the data of other investigators. In the previously cited Chicago study (13), 86 foster children of mentally defective mothers, adopted under the age of 5 years, had an average IQ of 95.1. In a group studied by Speer (63, 64), the IQ's of 12 children of feebleminded mothers, placed in boarding homes under the age of 3, averaged 100.5. In the same study, 16 children who had remained with their feebleminded mothers until they were from 12 to 15 years old had an average IQ of only 53.1.

In a study conducted by Stippich (69) at the University of Minnesota, 48 children of feebleminded mothers were compared with 29 children of normal mothers, all 77 children having been placed in boarding homes or institutions before the age of one year. The last intelligence test given these children, at an average age of 4½ years, showed a mean IQ of 89.38 for the "experimental" group (with feebleminded mothers) and 103.63 for the control group. It should be noted, first, that this study is in general agreement with those pre-

[18] Originally reported as 87, but when corrected by using adult CA of 15 rather than 16, a mean of 93 was found (41).

viously cited in finding that, when the children of feebleminded mothers are reared in normal homes, their IQ's average considerably higher than the IQ's of their mothers. The mean IQ of the mothers in this group was about 61, with a range from 32 to 77. Among the children of these mothers, 15 had IQ's above 100, and only 6 below 70. The children undoubtedly included some, at the lower end of the distribution, whose feeblemindedness resulted from unidentified structural deficiencies which limited their behavioral development. Such structural deficiencies, of course, could themselves be determined either by hereditary factors or by prenatal or natal environmental factors.

Certain other findings in the study by Stippich are of interest. Most of the children in both experimental and control groups showed a *rise* in IQ from early to later tests. These data thus lend no support to the argument that the children of feebleminded mothers may "fall behind" as they grow older and are measured with more reliable tests of intelligence. That both groups made a poorer showing than most foster groups is understandable when we consider that none of these children were adopted, but rather that they were placed in boarding homes. Such homes are generally of lower socio-economic level than foster homes, the children often being taken for boarding in order to supplement the family income. The interest and attention shown toward the child are usually less than in the case of an adopted child. Moreover, most of the children were shifted about from home to home, or from institution to home, the number of different placements per child ranging up to 9. Such a situation is not conducive to good adjustment or optimum intellectual development. As for the comparison between the final IQ's of experimental and control groups, it seems likely that some selective placement occurred and may account for part of the differences in IQ's. The author maintains that there was no tendency for the agencies to place children of feebleminded mothers in lower-level boarding homes. The distributions of occupational levels suggest, however, that some such tendency may have operated. For example, 14.6% of the experimental and only 8.2% of the control placements were in homes classified in the next to the lowest occupational category; 1.8% of the experimental and none of the control placements occurred in homes in the lowest occupational category. Moreover, it is possible that in other aspects of the home environment, not indicated by the crude occupational cate-

gories, the boarding homes in the experimental group may have been inferior. Thus even within the same occupational category the boarding homes in the experimental group may have represented the lower end of the scale.[19]

In criticism of some of the Iowa and Chicago results, it has been pointed out (41) that the mothers may not have been truly feebleminded, but that their IQ's may have been spuriously lowered because of testing conditions. If, for example, the mother is tested shortly before or after the birth of an illegitimate child, her emotional condition may not be conducive to good performance on an intelligence test. On the other hand, the results were no different when only mothers who had been institutionalized as feebleminded were included (63, 64, 69). Interpretation of the results of any of these studies is difficult, however, without information regarding the father's intellectual status. It can be argued that the child may have "normal heredity" if his father is normal, even though the mother was defective. Still another possible explanation is that the mother's feeblemindedness resulted from either prenatal or postnatal environmental factors, and that no hereditary factor—which could be transmitted to the child—was involved.

Concluding Evaluation of Research on Foster Children. In evaluating the contribution which the study of foster children as a whole has made to the analysis of heredity and environment, four major points merit consideration. First, all investigators agree in finding that intellectual development is affected, to a greater or lesser degree, by the type of home environment in which the child is reared. Secondly, the existing conditions of adoption make a more precise analysis of contributing factors impossible. There are too many unknown or uncontrolled variables whose influence cannot be isolated. Thirdly, the study of foster children is not—as has frequently been implied—a technique for comparing the relative contribution of "heredity" and "environment." It is at best only a means of investigating the influence of *one phase* of environment, namely, the type of home in which the individual has lived for a certain number of years (often a rather small number!). Other important aspects of the environment are not

[19] That the home environments of the two groups were not fully comparable seems quite clear. The experimental group included, for example, one child who had spent the first nine months of his life with his feebleminded mother and the rest of the time in an institution for the feebleminded, never having been in a boarding home at all!

covered. Schooling, for example, is relatively uniform for all individuals in the group. It would thus serve as an equalizing influence, tending to reduce the effects of varying home environments. Other influences outside the home, including community groups, organizations such as the Boy Scouts, and the like, are also probable equalizing factors. Prenatal and natal conditions are another set of environmental influences which are not considered in these studies. It would thus be misleading to regard the foster children studies as indicating the contribution of "the environment" to individual differences. They can only show the relative contribution of one restricted aspect of environment, as against all other combined influences of environment as well as heredity.

Finally, because of placement policies and practices, even that phase of environment which is investigated, viz., parental and home status, is artificially restricted. If the total range of American homes were covered, reaching down to the most deficient, then the observed effect of home environment upon intellectual development would probably be greater. Extending the range still further, to include other cultures where the standard of living is lower, would increase the relative contribution of home environment even more. It is also doubtful whether those aspects of the child's home environment which are most important for intellectual development have been adequately covered by the ratings of home environment which have been employed. An index which concentrated exclusively upon the most relevant characteristics might yield a higher correlation with IQ.

INSTITUTIONAL ENVIRONMENTS

Closely related to the analysis of foster family relationships is the study of children reared in institutions. Despite the apparent uniformity of their institutional home, such children generally show nearly as wide individual differences in intelligence as children living in their own homes. Moreover, in one investigation conducted in England (32), correlations in the .20's and .30's were found between the intelligence test scores of orphanage children and the occupational status of their own fathers. It should be noted that these children were placed by the institution in boarding homes until the age of 6. From 6 to 16 they lived at the orphanage, where they attended the same school. Since the occupations of the fathers were known to the

orphanage staff, one wonders to what extent selective placement and selective treatment in the institution may have artificially raised the reported correlations. It was also found that, among children admitted to the orphanage before the age of 3, intelligence test scores showed a lower correlation with parental occupational level than in the case of children who remained with their parents after the age of 3. In another British study (28) on orphanage children aged 9 to 16, a similar relationship between child's intelligence and parental occupation was noted. But the intellectual differentiation between occupational classes, as well as the extent of individual differences within any one class, tended to decrease as length of institutional residence increased.

A fairly well-established finding is that orphanage children on the whole have lower IQ's than those reared in either boarding homes or foster homes (12, 17, 36, 57, 78). In itself, such a finding permits of at least two explanations. First, *selective factors* may gradually eliminate the brighter children from an orphanage group, since such children are the most likely to be chosen for adoption. Secondly, *institutional environments* in general are relatively unstimulating to the developing child. Orphanages vary widely among themselves, of course, in the type of environment which they provide. Problems of overcrowding, staffing, space, equipment, and other facilities naturally produce differences in the amount and type of stimulation which the child receives. The ratio of adult staff members to children varies in different orphanages from about 1:2 to about 1:25 (cf. 78). To a certain extent, these differences in institutional environments are reflected in the IQ's of the children. Some of the apparently inconsistent results found by investigators in different orphanages are probably attributable in part to such institutional differences.

In a study of infants between the ages of 6 and 12 weeks, Gilliland (17) compared the performance of over 300 institutional infants with an equal number of infants living in their own homes. The IQ's of the institutional infants on the Northwestern Infant Intelligence Test averaged significantly lower than those of the infants in private homes.[20] Of the 40 items in the test, 18 showed a significant difference in favor of the infants in private homes. These items dealt with behavior which would be influenced by the nature and extent of the child's contacts

[20] The difference between these means was significant at a high level of confidence, the critical ratio being over 4

with his social and physical environment. Items concerned primarily with maturational changes showed no difference between institutional and non-institutional groups.

Several of the Iowa investigations were concerned with orphanage children. In one orphanage which offered relatively little opportunity for intellectual development, a nursery school was set up for a trial period (58). Some evidence was found that the IQ's of the nursery children tended to rise, while those of the other orphanage children dropped somewhat. Unfortunately, long-range comparisons could not be made satisfactorily in this study, since children were eliminated at frequent intervals from both groups, owing to adoption, and others were admitted.

In another widely quoted Iowa study (56, 57), 13 orphanage children under 3 years of age, with IQ's ranging from 35 to 89, were placed as "guests" in an institution for feebleminded women, one or two children being placed in each ward. There were about 30 women in each ward, ranging in chronological age from 18 to 50 and in mental age from 6 to 12. Despite their own intellectual backwardness, these women had higher mental ages than the children, of course, and were thus able to provide considerable intellectual stimulation for them. Together with the ward attendants, the feebleminded women evidently lavished attention and affection upon their young visitors. After about 18 months of this regimen, the infants gained an average of 27.5 IQ points, while a control group which had remained in the orphanage lost an average of 26.2 points. The contrast between these two groups is undoubtedly exaggerated by the regression effect, which is considerable at these age levels owing to the unreliability of infant tests (cf. Ch. 8). The "experimental" group had an *initial* average IQ of 64.3, while that of the control group was 86.7. Through regression, the initially lower group would be expected to rise somewhat and the initially higher to drop, since they were both samples of the same orphanage population. It is unlikely, however, that the entire difference can be attributed to regression.

A word should be added regarding the apparent contradiction between the above findings and some of the previously cited results on children of feebleminded mothers. It will be recalled that those children who remained with their feebleminded mothers tended to lose in IQ with age, while those who were placed with more intelligent foster parents seemed to develop normally. If the specific stimulating

conditions in the two types of studies are considered, it will be seen that the situations are only superficially similar and that no real contradiction exists. A home conducted by a feebleminded mother is probably poorly organized and inefficiently run. The mother herself, finding it difficult to cope with the everyday problems of living, may have little time or energy left to devote to the child. The emotional atmosphere, too, may be unfavorable in many such homes because of frustrations, economic difficulties, irresponsibility, and similar conditions. The institutionalized feebleminded girls, on the other hand, were free from other responsibilities, had little else to keep them occupied, and enthusiastically welcomed the diversion of caring for and entertaining the one or two babies in their ward. The amount of adult attention which the child receives would thus be quite different in the two situations.

The important part which attention from adults plays in the intellectual and emotional development of the child is being increasingly recognized. An interesting demonstration of this fact was provided by a comparative study of the development of infants in two European institutions (65). One of the two, described as "Foundling Home," was an ordinary orphanage in which hygienic and medical care was excellent, but adult contacts and other forms of stimulation were at a minimum. The infants in Foundling Home were kept isolated in cots, with no toys or objects other than bedding and clothing. There was very little for the child to see and practically no opportunity for locomotion. Essential physical care was provided by a trained nursing staff, each nurse having charge of eight babies. The other institution, designated as "Nursery," was established to care for the new-born babies of delinquent girls in a penal institution. It was closely comparable to Foundling Home in its physical and medical facilities, but the Nursery children were less isolated, had a certain number of toys, and were cared for by their mothers under the supervision of the nursing staff. The investigator gives some evidence to show that in terms of parental background the children in Foundling Home were superior to those in Nursery.

Both groups were examined periodically with the Hetzer and Wolf Baby Tests. The initial developmental quotient, based upon performance during the first four months of life, averaged 101.5 for the 69 Nursery children, and 124 for the 61 children observed in Foundling Home. By the end of the first year, the developmental quotient of the

Foundling Home group had dropped to 72, while that of the Nursery group averaged 105. Subsequent observations showed that the Nursery quotient remained close to normal, while that of the Foundling Home children continued to drop, reaching an average of 45 by the end of the second year. It should be noted that while the Foundling Home environment represented an extreme degree of isolation and lack of stimulation, the Nursery group received more attention, on the whole, than children in the typical family situation. For the mothers in the delinquent institution, the care of their child was one of the few sources of satisfaction and pride. The large majority of these children thus received an excessive amount of adult attention.

A considerable proportion of the infants in Foundling Home developed what appeared to be a clear-cut clinical syndrome, including extreme depression, retardation in all behavior development, and—in severe cases—complete withdrawal and immobility (65, 66). In its early stages, this condition could be improved by returning the mother to the infant or, if that was impossible, by placing the child where he was free to move about and had contact with other children and adults. If, however, the depressed condition was allowed to continue unchecked for about three months or more, the child failed to respond to such changes in treatment, the damage to its behavior development appearing to be permanent. A combination of emotional and intellectual deprivations is apparently involved in the extreme conditions represented in these observations.

Psychiatric disturbances among children reared in an institutional environment during the first year of life have also been reported by a number of other investigators.[21] Ribble (50) has repeatedly called attention to two general types of reaction which commonly develop among institutionalized infants who have received insufficient "psychological mothering." Some show a sort of *negativism*, which may include loss of appetite, failure to assimilate food, muscular tension and rigidity, and violent screaming. The other type of reaction, which Ribble characterizes as *regression*, consists of excessive depression, quiescence, and inactivity amounting almost to stupor. This condition often leads to a "wasting away" despite adequate diet and physical care.

Because of methodological difficulties, small number of cases, and

[21] For reference to a number of these studies, cf. Spitz (65), Goldfarb (18), and Ribble (50).

similar limitations, the studies on institutional environments can do little more at this stage than provide promising leads for future research. The bulk of their evidence does, however, point in the same direction as the results of other types of investigations. All such evidence indicates that certain aspects of the child's home environment may exert considerable influence upon his subsequent behavior. A close relationship between the child and one or more adults appears to be an important prerequisite to both emotional adjustment and intellectual development. The data on language development, reported in the section on twins, should also be recalled in this connection. Similarly, a comparison of the language development of preschool children living in their own homes with that of preschool orphanage children favored the former group (43). The children reared in their own homes excelled in size of vocabulary and in the variety of subjects about which they talked. These differences persisted when sex, chronological age, and mental age were held constant. Among the environmental factors mentioned to account for the observed differences are the adult-child ratio and the number and variety of experiences about which the children can talk.

CHAPTER

12

Bodily Conditions and Behavior

REPEATED REFERENCE HAS BEEN MADE in the preceding chapters to the possible limits of behavior development set by the individual's structural characteristics. Age changes in behavior provided some illustrations of such limitations. Until the infant has attained a prerequisite level of sensory, neural, and muscular development, for example, certain specific behavior functions may be very difficult or impossible to learn. Similarly, the physical deterioration in senescence is likely to curtail many of the person's activities. Physical differences among individuals may likewise contribute to the observed individual differences in behavior. It is apparent that extreme sensory defects, for example, can so seriously handicap the individual that even special training may not bring him up to a normal level of performance. Other parts of the reacting organism may, through either their deficiency or their superiority, affect the development of psychological traits. Certain forms of feeblemindedness are undoubtedly traceable to structural defects which prevent the attainment of a normal level of behavior development, despite adequate stimulation.

It should be clearly recognized, however, that any structural characteristic whose influence on behavior development is demonstrated serves as a *necessary but not a sufficient condition* in such development. In other words, the presence of the prerequisite physical factor does not in itself determine behavior, but simply makes a certain kind of behavior development possible, *if the proper stimulation is available.* Moreover, structurally imposed limitations are probably less effective than is commonly supposed, since individuals rarely attain the degree of development set by their physical capacity. Physiologi-

cal and biological conditions may thus be regarded as "participating factors" in psychological reactions, rather than as the underlying determinants of any behavior function.[1]

The relationship between structural and psychological characteristics is of special interest for an analysis of the contribution of hereditary factors to behavior. In so far as hereditary factors affect behavior, they must do so through their control of structural development. Obviously the individual does not "inherit" functions as such. The study of any correspondences between behavior differences and differences in bodily condition would thus seem to be a necessary first step for a realistic consideration of the role of heredity in behavior.

There are a number of well-known types of intellectual and emotional disorders which are directly traceable to extreme glandular malfunctioning, the deterioration of tissues, the effects of drugs or infections on the nervous system, and similar abnormal conditions. The behavior symptoms associated with paresis, delirium tremens, or cretinism, for example, can be clearly related to the physical effects of syphilitic infection, alcohol, or thyroid deficiency, respectively. Many other similar illustrations could readily be cited. The present discussion, however, is not concerned with these extreme and pathological conditions. The question before us now is essentially this: "To what extent are individual differences in behavior associated with the structural differences commonly found within the normal range of variation?"

CRANIAL AND CEREBRAL MEASUREMENTS

Popular interest in the size and shape of the skull was considerably stimulated by the pseudo-science of *phrenology*, initiated by Gall in the last years of the eighteenth century. Phrenology was based upon a false notion of the functions of the various parts of the cerebral cortex. The phrenologists maintained that each area of the brain controlled a particular intellectual or moral function, such as mechanical ingenuity, veneration, domestic impulses, and other equally complex and vaguely defined activities. They asserted further that the over- or underdevelopment of such behavior characteristics could be diagnosed by examining the protrusions of the skull. The location of a particular

[1] For a searching and systematic analysis of the role of structural factors in psychological functioning, cf. Kantor (40).

"bump" was taken to mean that the function allegedly controlled by the corresponding cortical area was highly developed in the given individual.

It would seem unnecessary to refute such an obviously untenable doctrine were it not for its enduring popularity among the general public and its lucrative practice by a considerable number of charlatans. In the first place, phrenology is founded upon the erroneous assumption that there is a close correspondence between the shape of the skull and that of the brain. Such a correspondence is hardly to be expected, in view of the cerebro-spinal fluid and the several layers of membrane which intervene between brain and skull. It should also be noted that size does not provide a satisfactory index of degree of development within the nervous system. It is the complexity of interrelation of the minute nerve cells and other microscopical characteristics of nerve matter that are probably related to efficiency of function. Moreover, the type of trait which phrenologists ascribe to different brain areas is quite unlike the functions discovered through investigations of cortical localization. Connections have been demonstrated between certain muscle groups or sense organs and specific brain areas, but this is a far cry from the localization of "literary propensities" or "love of dumb animals" on the cortex!

Phrenologists have also tried to show that cranial capacity as a whole, or total brain size, is related to intelligence. Their evidence for this, as for their other assertions, is based upon selected examples and is therefore worthless. It is true, for example, that a certain type of idiot—the microcephalic—has a very small skull. But there are also idiots with normal or very large skulls. A few men of genius may be found with very large brains,[2] but others are likewise found with smaller than average brains. The question can be settled only by precise measurement of large numbers of unselected cases.

Investigations on the relationship between cranial capacity and intellectual achievement have generally yielded negative results. In a number of early studies in which average cranial dimensions of bright and dull groups were compared, the data are ambiguous and difficult to interpret.[3] The differences between the averages are always extremely small and occasionally inconsistent from one comparison to another. In certain investigations the measures taken on the living

[2] A favorite example is Daniel Webster, whose head circumference measured 24½ inches.

[3] For a survey of these data cf. Paterson (65, Ch. 3).

skull were not good indices of brain capacity. The groups employed often varied widely in age. When children are included, this may produce a spurious relationship between size of head and intelligence, since the older subjects will have larger heads and at the same time will obtain higher scores on intelligence tests. Finally, the estimates of intelligence were frequently crude and unreliable.

The first well-controlled study on cranial measurement in which adequate correlational analysis was employed is that of Pearson (66). Measures of head length, head breadth, and cephalic index [4] were obtained on three groups, including 1010 Cambridge University students, over 2200 12-year-old school boys, and over 2100 12-year-old school girls.[5] It will be noted that age was held constant among the children by selecting only 12-year-olds. The subjects were classified into intellectual levels on the basis of teachers' ratings and scholastic records. The correlations between intellectual level and cephalic index were −.06, −.04, and .07 among the university students, school boys, and school girls, respectively. For length of head, the correlations in these three groups were .11, .14, and .08, and for breadth of head .10, .11, and .11. These correlations speak for themselves, being too low in every case to indicate any appreciable trend. The very low and inconsistent correlations with cephalic index lend no support to a frequently proposed theory that the "long-headed" individuals (with a low cephalic index) are the more intelligent. Nor do they support the opposite view, also occasionally voiced, that the "broad-headed" (with high cephalic index) are the more intelligent.

More recent investigations by the correlation method have in general substantiated Pearson's findings. Murdock and Sullivan (62) report a correlation of .22 between head diameter (obtained by averaging maximum head width and maximum head length) and IQ [6] with about 596 elementary and high school pupils. By the use of IQ's and by the conversion of physical measurements into deviations

[4] Cephalic Index $= \frac{100 \times \text{head width}}{\text{head length}}$. Length of head is measured from the space between the eyebrows to the farthest projection at the back of the head; head width, or breadth, is the distance from left to right sides, measured from the points of maximal protrusion above each ear. The following is a common classification of cephalic index:

Dolichocephalic, or long headed	CI below 75
Mesocephalic, or medium-headed	CI between 75 and 80
Brachycephalic, or broad-headed	CI above 80

[5] The number of cases differed slightly for each measure.

[6] Found from a number of group intelligence tests.

from the average of each age-sex group, the influence of age was held constant. In a study of 449 medical students in Scotland, Reid and Mulligan (75) found a correlation of .08 between cranial capacity and scholastic achievement. Cranial capacity was calculated by taking the product of length, breadth, and width of the head, with allowance for thickness of different parts of the cranium. Scholastic achievement was determined by performance on standardized examinations in three courses which were taken by all the students.

Sommerville (92) obtained correlations of .10, .03, and .09 between the scores of 100 male college students on the Thorndike Intelligence Examination for High School Graduates and measures of head length, head width, and head height, respectively. The correlations were no higher between intelligence test scores and cranial capacity as estimated from the three given head dimensions. Employing one standard formula for the computation of cranial capacity, Sommerville found a correlation of .11 with intelligence test scores; with another formula, the correlation was .10. These findings were closely confirmed in a more recent investigation by Broom (9). Cubic brain capacity, as estimated from external measures, gave insignificant correlations with intelligence test scores in a group of 100 college men and 100 college women.

It thus seems to be quite conclusively established that no appreciable relationship exists between intellectual level and either cranial capacity or head shape as determined by the cephalic index. The correlations, although generally positive, are so low as to be of doubtful significance. Some dissenting views are occasionally expressed, even by modern writers, advocating the use of cranial measurement in the diagnosis of intellectual development.[7] But their evidence is ambiguous and their arguments are weak and inconsistent. Additional data on more detailed cranial conformation will be discussed in a later section of the present chapter, in conjunction with facial measurements.

In recent years, interest has shifted from the gross characteristics of brain size and shape to less obvious conditions which are more likely to influence brain functioning (47, 84). For example, the *thickness of the cerebral cortex* has been measured, and the *concentration and distribution of cells* has been determined by taking sample counts in different sections. *Fissurization*, or the nature and extent

[7] Cf., e.g., Porteus (73) and, for a critique of the data, Paterson (65).

of "folds" in the cerebral cortex, has been considered to be significant by some writers, since the brains of lower animals and immature organisms are relatively smooth. Family similarities have also been noted in fissurization. Attempts to discover any correlation between behavior characteristics and any of these brain conditions, however, have so far met with consistent failure. Similarly, the *chemical composition* of the brain, although undoubtedly an important factor in certain pathological cases, has as yet never been related to specific forms of behavior within the normal range of variation. The present state of knowledge regarding the relations between these various brain conditions and behavior has been aptly characterized by Lashley. In a paper presented in 1947, he states (47, p. 326): "An attempt to relate phylogenetic and individual differences in behavior to brain structure is therefore rather an adventure in correlating the mysterious with the unknown."

A relatively new and promising field of brain research centers around the *electroencephalogram* (EEG), a record of the minute changes in electrical potential generated in the brain (16, 50). A special advantage of this technique is that it permits a study of brain function in the living organism. Since many important properties of tissues are lost upon death, postmortem brain examinations may exclude essential facts. By means of electrodes attached to the scalp, the minute "brain waves," or fluctuations of electrical potential in the brain of the living person, are picked up, magnified, and recorded graphically. It will be recalled (Ch. 5) that evidence of electrical activity in the cortex of the guinea pig was found during fetal life, although in the human brain no conclusive evidence of such activity has been found until some time after birth. Several types of rhythmic changes in electrical potential, differing in frequency and amplitude, have been identified in the adult human brain. Some of the most clear-cut results have been obtained with the *alpha waves,* which have an average frequency of about 10 per second and are found in normal children and adults during a relaxed waking state. Fairly consistent age differences have been observed in the frequency and amplitude of these alpha waves, as well as in the per cent of time that the alpha rhythm is present (49). Between the ages of 3 and 10, for example, there is a progressive increase in the frequency of the alpha rhythm. Sufficient data have been gathered to establish age norms in various aspects of the EEG (49, 50). Individual differences

in EEG have also been observed, the individual characteristics being maintained with considerable consistency on successive retests (49, 50).

Such findings have led certain investigators to inquire whether the developmental changes in EEG are related primarily to chronological age or to mental age. In studies on several types of feebleminded adults, Kreezer (44) reported a number of significant but generally low correlations between mental age and certain characteristics of the alpha waves. Several points should be noted, however, in interpreting these results. First, the groups studied were usually small and many of the correlations were barely significant. Secondly, different characteristics of the alpha waves yielded significant correlations in different types of feeblemindedness, suggesting that whatever relationship exists is certainly not a simple one. Thirdly, the significant relationships were confined to types of feeblemindedness having other clearly recognizable physical deficiencies. They were not substantiated in a group of "undifferentiated" feebleminded cases with no observable physical pathology. This suggests that the disturbance in EEG may be associated with the other pathological physical conditions and need have no implications within the normal range of variation. Other investigators have also failed to discover any significant relationship between EEG characteristics and intellectual level among undifferentiated feebleminded subjects (16, 49).

What little direct evidence is available on normal subjects is also negative. In one study on normal children (43), for example, a significant correlation of .50 was found between alpha frequency and IQ among 48 8-year-olds, but an insignificant correlation of .12 was obtained among 42 12-year-olds. It is possible that, among the younger children, individual differences in the level of physical development within a single year of chronological age may account for the significant correlation. In a group of 1100 aircrew candidates between the ages of 18 and 33, no relationship was found between intelligence test score and alpha frequency (82).

In the area of personality characteristics and emotional abnormalities, EEG results are scanty and inconclusive, with the exception of the extensive body of data on epilepsy. It has been quite clearly established that epileptics show characteristic deviations in EEG, and that relatives of epileptics who have not themselves developed any of the clinical symptoms of epilepsy show similar disorders in EEG

(48, 50). It is also interesting to note that numerous case reports of children with behavior disorders show abnormalities of the EEG, some of them of an epileptoid form (50). We can probably attach little significance to the fact that several attempts to correlate personality test scores and EEG characteristics in adults as well as children have yielded inconsistent and inconclusive results (32, 49). In such cases, the inadequacy of the personality tests as measures of behavior characteristics may have been partly responsible for the negative findings.

PHYSIOGNOMY AND RELATED SYSTEMS

There are many firmly entrenched popular beliefs regarding the "meaning" of various facial and other bodily characteristics. The high forehead of the intellectual "high-brow," the shifty gaze of deceitfulness, the firm chin and square jaw of determination, the tapering fingers of the artist, and a host of other traditional associations which the reader can easily name have found their way not only into poetry and fiction but also into the snap judgments and "hunches" of everyday life. Similarly, we frequently hear of alleged personality differences between blondes, brunettes, and redheads, between blue-eyed and brown-eyed persons, or between those with a "convex" and those with a "concave" profile. Many of these beliefs can be traced to ancient times. During the last quarter of the eighteenth century, a number of them were organized by Lavater into a system of character analysis known as "*physiognomy*." Today, this system is about as popular among charlatans as is phrenology—and equally unfounded.

A series of carefully controlled investigations designed to check many of the assertions of physiognomy were conducted under the general direction of Hull (37). The relationship between convexity of profile and several personality traits, which is often stressed by self-styled "experts" in physiognomy, was studied by Evans (19). The subjects were 25 college women, all of whom were members of the same sorority. Such a group was chosen because of their close acquaintance with each other and their consequent ability to rate each other with a fair degree of accuracy. For the same reason, all individuals who had not been members long enough to be well known

were excluded from the study. Each girl ranked the remaining 24 in six personality traits, including optimism, activity, ambition, will power, domination, and popularity. The average or consensus rank of all 24 judges for each girl was computed as a final estimate of each trait. The subjects were also rated in a similar way for degree of blondness. A specially devised mechanical instrument was employed to read off directly the "angle of convexity" of the profile. In order not to exclude any possibilities, convexity was measured in five different ways, such as whole face, upper face only, convexity without including the nose, and so on. Height of forehead was also measured.

The correlations between each of the measures of convexity of profile or height of forehead and each of the six personality traits were low and often inconsistent with expectation—that is, a correlation which would have been expected to be negative on the basis of the physiognomists' claims was often positive, and vice versa. The highest correlations were a +.39 between "convexity of whole face with nose omitted" and "activity" rank, and a −.39 between height of forehead and "will power" rank. Even these correlations, however, are not significant in view of the small number of subjects, and could have resulted from chance errors of sampling. The correlations for blondness ranged from +.28 with will power to −.26 with optimism. These are also too low to be significant.

A further point to bear in mind in evaluating these correlations is that the existence of a widespread bias among the judges regarding the association of facial and personality characteristics might in itself produce a correlation. Since tests were not available for the traits under consideration, it was necessary to resort to associates' judgments; but this procedure is inconclusive when widespread popular beliefs are present.

Facial and cranial measurements were combined in a study by Sherman (83). A group of 78 freshmen in an engineering college were measured by means of a specially designed "radiometer." A total of 15 distances and 4 angles were found for each subject, and each of these measures was then correlated with academic grades. The correlations with the combined grades on all courses ranged from −.26 to +.34. It is interesting to note that height of forehead correlated −.15 with academic grades. This corroborates the low negative correlations found by Evans between height of forehead and

several personality traits which might be expected to manifest themselves in school work. If such a tendency were established, it would indicate a reversal of the popular notion of a "high-brow"!

Similar studies have been conducted by a number of other investigators. One study had as its object to determine whether there is any relationship between the shape of the hand and a number of traits suggested by "chirognomists" (cf. 37, pp. 145–146). The results were clearly negative. Numerous experiments have been conducted to discover whether it is possible to judge intellectual or emotional traits from photographs, as might be expected if these traits were manifested in facial characteristics. All these investigations showed a lack of correspondence between the various physical characteristics and the behavior traits with which they were allegedly associated. There were cases, however, in which the judges agreed rather closely among themselves, a finding which suggests the prevalence of such popular stereotypes, or conventionalized physiognomic symbolism.

It should be pointed out in conclusion that even when significant correlations are found between certain facial or cranial characteristics and psychological traits, as in the case of a few of Sherman's measures, the correlations are still too low to give any information about *individuals*. They simply indicate a general trend in the group which may result from a few extreme cases. In so far as the correlation is far below 1.00, it shows that there are many individual exceptions to the general trend. The presence of these exceptions or reversals of relationship proves that whatever direct influence any such physical factor may exert upon behavioral development is very weak and can easily be obscured by other, more potent factors.

A further point to note is that as long as a certain belief is widely prevalent regarding the association of a given physical characteristic with an intellectual or emotional trait, this may in itself influence the individual's development. If a person is commonly mistrusted by his associates and is not given any responsibility, it is difficult for him to be open and sincere. If a child is regarded as dull and stupid, he may easily come to believe it himself and act accordingly. Moreover, the school child whose appearance fits the popular stereotype of "stupidity" will probably receive poorer school grades than his "bright-looking" classmate whose actual achievement may be no better. The social and motivational influence of a widespread prejudice cannot be ignored. A vicious circle is initiated by such a situation: the

more widespread the prejudice, the more effective it will be and the more evidence can therefore be found which seems to support it.

From these considerations it is apparent that any relationship which may exist between facial characteristics and psychological traits cannot be large. Even the slight correlation occasionally found is far from conclusively established because of many remaining uncontrolled factors. Should a slight correspondence be proved between certain facial or cranial conformations and behavior, such an association could result from a common dependence of both types of characteristics upon the same underlying condition. The activity of the endocrine glands offers possibilities for such a connection. In certain extreme pathological cases as, for example, thyroid deficiency, the resulting condition includes typical physical as well as mental symptoms. It is barely possible that certain facial characteristics, as well as emotional or intellectual traits, are influenced within their normal range of variation by over- or underactivity of some endocrine gland. This, of course, is only speculation. The field of endocrinology is far too complex and too young to offer any clear-cut answers to such a query.

BODILY DIMENSIONS

Gross bodily dimensions, proportion of trunk and limbs, height in relation to weight, and similar structural characteristics have also been suggested as possible indices of intellectual or emotional status. Since much of the material in this field has been collected to test out the various "type theories" proposed from time to time, the discussion in this section will be supplemented in the following chapter. Only the data on gross size and absolute measures will be treated here, the material on relative proportions and body type being reserved for Chapter 13.

Similarly, we are not concerned with gross malformations and pathological conditions. Many of these conditions, familiar to anyone who has seen circus "freaks," have been definitely traced to glandular disorders. Thus gigantism, a condition in which the individual may attain a height of seven or eight feet,[5] results from oversecretion of a pituitary hormone. Dwarfism, or stunted growth with normal bodily

[5] The "giant" with the Ringling Bros. Barnum & Bailey circus was reported to be 8 feet, 6½ inches tall

proportions, is produced by insufficient pituitary secretion. No definite intellectual defect has been demonstrated in these cases. Cretinism, associated with an underactive thyroid, is characterized by abnormal bodily development and proportions as well as by intellectual defect, sluggishness, and other behavioral disturbances. If we exclude cases which manifest obvious glandular dysfunctions or other pathological conditions, we still find a wide range in height and weight within the general population. It is into the relationships of these variations with behavioral characteristics that we now wish to inquire.

As in the case of cranial measurements, interest in body build has long been manifested. The search for a possible relation between body dimensions and intellect probably received a strong impetus from the popular view that the intellectually gifted were deficient in other respects. In particular, it was maintained that such individuals were weak, puny, and physically inferior. This notion of compensation was cherished widely because of its consoling character—it was no doubt accepted as the device of a benevolent nature to "even things up." In the effort to overthrow these unfounded beliefs, early research workers swung to the opposite extreme and asserted that the intellectually ablest were also the physically ablest and that a close correspondence exists between physique and mental ability.

Galton (22), for example, maintained that the number of physically superior individuals among his groups of eminent men (cf. Ch. 10) was greater than in the general population. Many studies on large groups of children have subsequently appeared which relied upon the comparison of averages for their conclusions.[9] Such investigations agree in finding a slightly higher average height and weight among the intellectually superior groups than among the normal, and slightly higher among the normal than among the dull. Intelligence was usually estimated quite crudely from school progress or teachers' ratings. The differences in averages were always so slight and the overlapping of groups so large that the degree of correlation between height or weight and intelligence would necessarily be negligible.

Investigations on the physical status of the feebleminded or the intellectually gifted child have yielded results which are equally difficult to interpret. When averages are compared, the feebleminded appear to be definitely below the norms in height and weight, and the

[9] For a summary of the early literature, see Paterson (65).

bright children above the norms. In Terman's extensive investigation (98) on gifted children,[10] a slight tendency was noted for the subjects to be above the age norms for American-born children in height and weight. L. S. Hollingworth (35) compared the heights of three groups, each composed of 45 children between the ages of 9 and 11. In the "superior" group were only children whose IQ's were above 135 (median IQ = 151); in the "normal," those with IQ's between 90 and 110 (median IQ = 100); and in the "inferior," those with IQ's below 65 (median IQ = 43). The subjects in the three groups were carefully equated, each child in the one group being "matched" with a child in the other two groups in respect to age, sex, and racial background, so that the influence of these factors was ruled out. In Table 17 will be found a frequency distribution showing the number of children in each group who fell within successive class-intervals in height, as well as the average height of each group.

TABLE 17 *Distributions and Averages of Height in Intellectually Superior, Normal, and Inferior Groups*

(From L. S. Hollingworth, 35, p. 80)

	Frequencies		
Height in Inches	Group A (Median IQ = 151)	Group B (Median IQ = 100)	Group C (Median IQ = 43)
55–59	12	2	1
50–54	30	30	18
45–49	3	13	23
40–44	0	0	3
Average height	52.9	51.2	49.6

Norsworthy (64), in an early but comprehensive survey of the characteristics of the feebleminded, obtained measures of height and weight on 157 mental defectives in special classes and in various institutions. She found the same slight differences in averages, with marked *overlapping*, 44% of the mentally defective children exceeding the median of normal children in weight and 45% in height.[11]

[10] Cf. Chapter 17 for fuller report.

[11] Complete overlapping would have been indicated if 50% of the feebleminded group had exceeded the normal median.

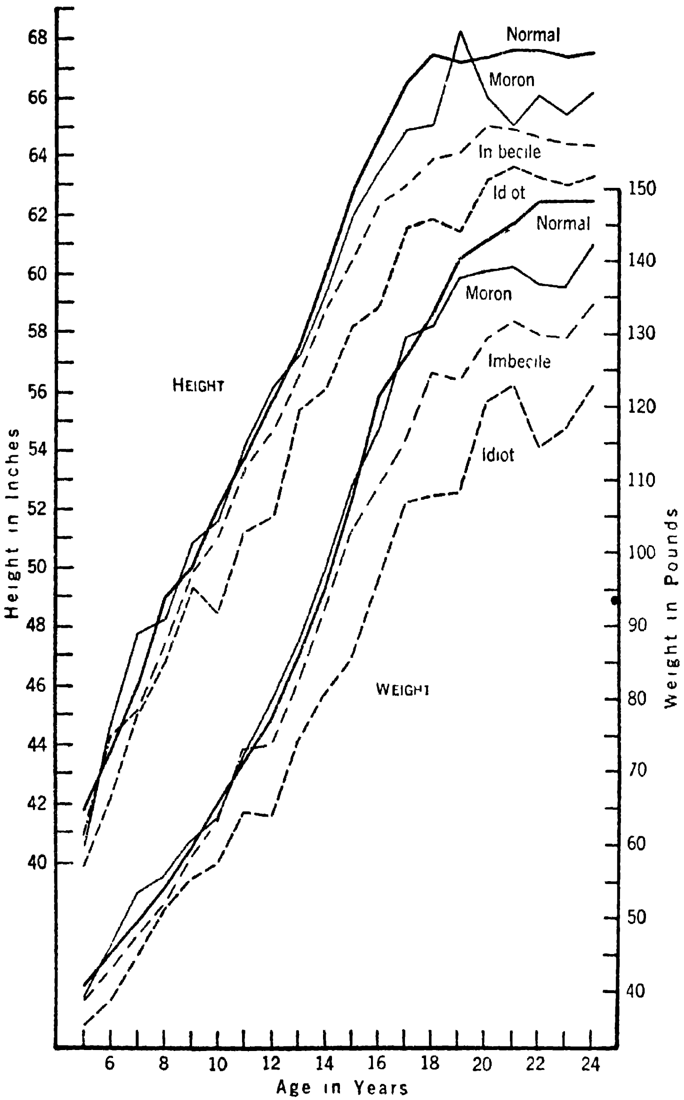


Fig. 68. Average Height and Weight of Feebleminded and Normal Boys at Successive Ages. (From Goddard, 24, p. 228.)

Goddard (24) collected extensive data on the height and weight of about 11,000 mentally defective individuals, ranging in age from early infancy to 60 years, in 19 American institutions for the feeble-minded. In Figures 68 and 69 are reproduced curves showing the

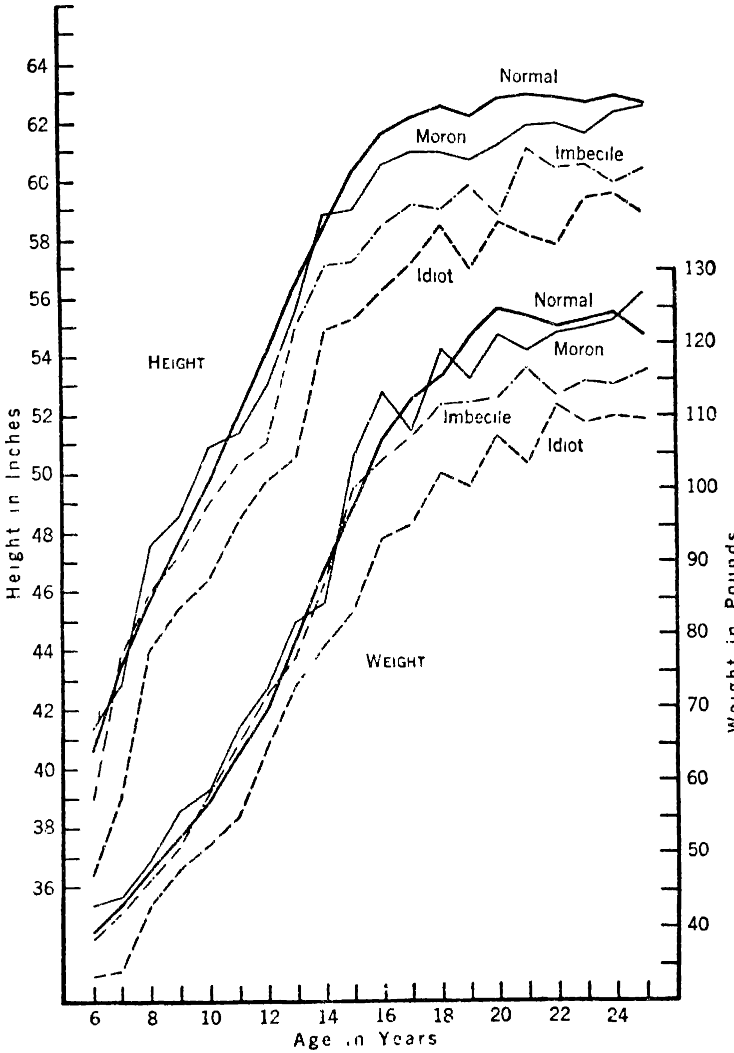


Fig. 69. Average Height and Weight of Feebleminded and Normal Girls at Successive Ages. (From Goddard, 24, p. 229.)

average height and weight of successive age groups within four intellectual levels; the data on boys are given in Figure 68, those on girls in Figure 69. It will be noted that the curves of the four intelligence groups exhibit a slight but consistent tendency for physical inferiority to parallel intellectual inferiority. The relationship is clearer during the adolescent and adult years than it is at earlier years.

Several factors may enter in to complicate the analysis of institutional data on the feebleminded. In the first place, those individuals with physical as well as intellectual defects are more likely to be committed to an institution. The feebleminded person who is physically fit or superior is less likely to be sent to an institution at all and more likely to leave the institution after he has received several years of training. Such individuals will have a greater chance to succeed in a routine occupation requiring strength and a good physique, with a minimum of thought and planning. The operation of such a selective factor might explain the divergence of Goddard's height and weight curves with age. Since only institutional cases were tested, the inferiority at the upper ages could have resulted from the fact that the physically strongest and ablest had left the institution. In addition, the norms in terms of which these groups are evaluated may not be comparable at successive ages. Such norms are usually established on school children because of the latter's ready accessibility for measurement. The norms at higher ages are frequently derived from high school students, a distinctly select group in respect to the general population. Finally, in a survey extending down to low-grade feebleminded levels, it is likely that several cases presenting special conditions such as cretinism are included; this would further lower the average physical measurements of the feebleminded group.

A more direct answer is provided by studies which employ the correlation technique with normal adult groups. In such studies, the complicating influences of age and of special pathological conditions are avoided. Moreover, the correlation coefficient provides an index of the degree to which a relationship between physique and intelligence exists among all individuals in the group. Average differences between groups, especially when slight, may mean little if the overlapping between groups is large. Brooks (8), employing 1118 junior high school, normal school, and college students between the ages of 13 and 20, correlated measures of height and weight with perform-

ance on several standardized group intelligence tests. Since correlations were computed separately for the two sexes and for several age groups, the subjects were classified into 17 groups ranging in number of cases from 16 to 139. The height correlations ranged from $-.09$ to $+.26$; those for weight ranged from $-.31$ to $+.26$. In the study by Sommerville (92) on college freshmen, described in an earlier section,[12] a correlation of .16 was found between intelligence and standing height, .13 between intelligence and sitting height, and .10 between intelligence and weight. The majority of these correlations are positive but so low as to indicate little or no appreciable relationship between general bodily size and intellectual level, when selective factors and other irrelevant conditions are ruled out.[13]

PHYSIOLOGICAL CONDITIONS

Efforts to discover what association, if any, exists between various physiological conditions and behavior characteristics have followed two principal approaches. The first is a comparison of the *relative frequency* of a particular physiological condition in groups or individuals differing in known intellectual or personality characteristics. The procedure can, of course, be reversed by comparing the psychological characteristics of individuals chosen on the basis of physical condition. The second and somewhat more direct procedure is to see what are the psychological effects of *treatment* for a given physiological condition. Such "before-and-after" studies, when feasible, provide a more clear-cut analysis of causal relationships. The number of physiological factors which could be considered in relation to behavior is almost without limit. We shall discuss a few typical illustrations which are of more general interest.

General Health. Is general health related to intellectual level? Are miscellaneous physiological defects such as enlarged glands, defective breathing, dental caries, diseased tonsils, and other common health disorders any more prevalent among dull than among bright persons in the general population? A number of extensive school surveys on this question have been conducted from time to time, using either school achievement or intelligence test scores as indices

[12] Cf. section on Cranial and Cerebral Measurements.

[13] Correlations found among children will be discussed in the section on Developmental Relationships.

of ability level (2, 17, 41, 57, 78, 94, 100). Some have yielded negative or ambiguous results. Among those showing some evidence for a relationship between frequency of defects and intelligence is the extensive study conducted by the U. S. Public Health Service on about 4500 school children in two counties in Illinois (41). The most relevant results of this survey are summarized in Table 18, which shows the relative frequency of each defect among children whose IQ's were under 90 and among those with IQ's of 110 or higher. Each entry in this table is a ratio of the number of defects in the designated IQ group to the number found within the normal IQ

TABLE 18 *Relative Frequency of Physical Defects among Bright and Dull School Children*

(Adapted from Kempf and Collins, 41, p. 1772)

Type of Defect	*Frequency of Defect*	
	IQ under 90	IQ 110 or over
One or more decayed teeth	111	82
Gingivitis	152	91
Defective tonsils	116	86
Adenoids	123	68
Other nasal obstructions	119	89
Enlarged glands:		
(1) Anterior cervical	117	81
(2) Posterior cervical	99	79
(3) Submaxillary	135	61
(4) Thyroid	114	94
Defective hearing (voice test)	157	72
Otitis media	147	45
Defective eardrum	174	69
Mastoidectomy (scar)	68	222
Defective vision (Snellen test)	128	90
Conjunctivitis	125	135
Strabismus	175	109
Speech defects	143	58
All heart defects	111	75
Nutrition: poor or very poor	106	81
Posture: poor or very poor	156	120
Fingernail biting	87	96
Evidences of rickets	110	115
Scoliosis	106	110
High arched palate	127	[illegible]7
Marked dental malocclusion	136	89
All skin diseases	178	136

range (90–100), the latter being taken as a standard. The first entry in the table, for example, means that the number of children with decayed teeth in the dull, normal, and bright groups was in the ratio of 111: 100: 82. The number in the middle group is always taken as 100, and the number in each of the other two groups is expressed as a ratio to 100. With only a few exceptions, there is a consistent tendency for each defect to be most common in the low IQ group and least common in the high IQ group. The investigators also report that the average number of different defects per child showed the same general relationship to IQ. The relationship, however, appears to be a general rather than a specific one. It is not any particular type of defect, but rather the presence of defects as such, which differentiates the high and low IQ groups.

Such group trends should be interpreted with considerable caution because of the many individual exceptions. A sizable proportion of dull children who are completely free from physical defects can be found in each survey; bright children with many defects can likewise be found. Correlations between intellectual or scholastic level and physical conditions have been uniformly low (cf. 65).

General health, as rated by a physician, also shows little or no appreciable relationship to intelligence within a typical sampling of school children. In one study (34) on 343 third and fourth grade American school children, those rated "good" in health had an average IQ of 104; those rated "poor" averaged 101. Moreover, those children whose health rating improved from the initial to the final examination showed no more gain in MA than those whose health became worse. Among intellectually backward children, on the other hand, the relationship between poor health or physical defects and IQ is much closer. In a survey of 14,379 retarded school children in Massachusetts, large and significant differences in average IQ were found between those having various physical defects and those free from defect (17). The mean IQ of the entire group, however, was 70.7, and over 500 children had IQ's below 50. The inclusion of feebleminded cases may account for the relationship in such a group.

Local Infections. It has sometimes been argued that any local infection in the body may, by releasing toxins into the blood stream, affect the functioning of the entire organism. Because of its superficial plausibility, such a theory meets with ready popular acceptance. At times it has led to exaggerated claims regarding the psychological

as well as physical improvement to be expected from the treatment of such conditions as diseased tonsils or infected teeth. The fact of the matter is, however, that no significant effects of any such conditions upon behavior characteristics have been demonstrated when proper experimental controls were observed. Most bodily mechanisms apparently have sufficient protection against any general effects which these local disorders may induce.

Among the extravagant assertions which have attracted popular attention from time to time is the claim that *dental caries* (decayed teeth) will interfere with a child's intellectual development. The evidence cited in support of such claims has never stood up under critical analysis. What few dependable data are available on this question indicate virtually no relationship between dental caries and intellectual level. Correlations between dental condition and intelligence have proved to be uniformly low and negligible (cf. 65, Ch. 6). Similarly, extensive dental treatment and prolonged training in oral hygiene were not accompanied by any greater gain in mental test performance than was found in a control group which did not receive these benefits (42).

Probably because of their relatively frequent occurrence among children of school age, *diseased tonsils* have also received their share of attention as a possible cause of intellectual backwardness. In the effort to test these claims, a carefully controlled investigation was carried out with 530 public school boys between the ages of 6 and 14 (77). All had been given the Stanford-Binet as a part of the regular school routine. On the basis of an examination by the school nurse or physician, the children were classified into two groups, the one composed of 236 boys whose tonsils were sufficiently diseased to require treatment, and the other of 294 boys whose tonsils were either not defective or so slightly defective as not to call for treatment. The average IQ's of the normal and defective groups proved to be 95.4 and 94.9, respectively. The percentage distributions of the two groups are given in Figure 70. The practically complete overlapping of these groups is apparent from an examination of the distribution curves.

As a further check on any possible influence of tonsillar condition upon mental development, 28 boys whose tonsils were subsequently removed were retested with the Stanford-Binet after a six months' interval. The gain in IQ made by this group was compared with that

of a control group of 28 boys who suffered from diseased tonsils but who had not been operated upon. The operated group made an average gain of 2.25 IQ points, as compared with an average gain of 3.28 in the control group. Finally, it was possible to test 21 subjects after an interval of from 10 to 17 months following their operation. The average gain in IQ made by this group was 3.0 points,

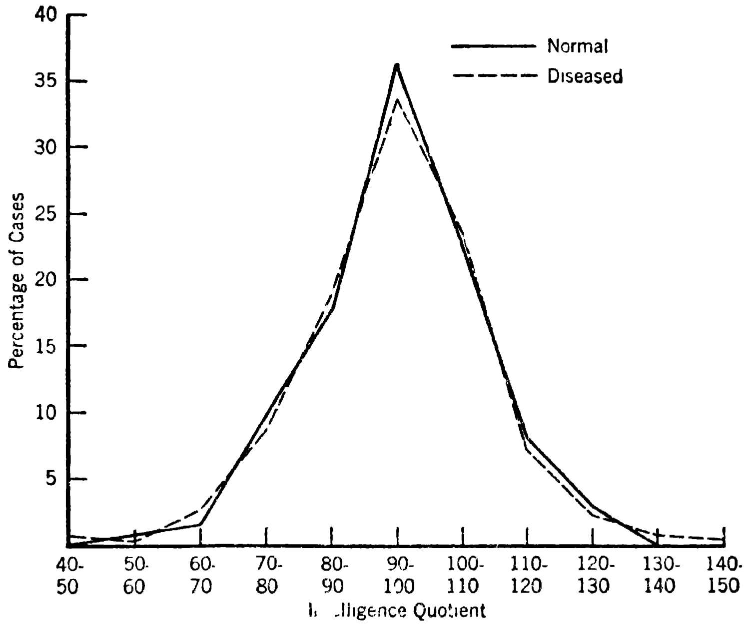


Fig. 70. Percentage Distribution of IQ's of Boys with Normal and with Diseased Tonsils. (Data from Rogers, 77 p. 29.)

while a control group of 31 cases gained 6.2 points. It is very doubtful whether further retests after a longer delay would reveal any effect of the tonsillectomy on intellectual development. The results of the two types of procedure followed in this study are thus mutually corroborative in demonstrating a lack of relationship between intellectual level and diseased tonsils. Not only was there no significant difference between the initial IQ's of normal and diseased groups, but also removal of the diseased tonsils produced no improvement in IQ which could be attributed to such treatment. These findings have been corroborated by more recent studies (51, 76).

Another popular belief is that *hookworm infection* produces mental defect, sluggishness, and apathy. Because of its prevalence among school children in certain parts of the country, this infection has attracted the serious notice of educators. Several studies have indicated the tendency for children with hookworm infection to be duller than those not so afflicted. In one of the most careful of these investigations (90), the Otis Intelligence Test was administered to 118 children in grades 3 to 7 of three rural schools located in a typical "hookworm area." Through medical tests, the degree of hookworm infestation was also determined for each child. Below are given the average IQ's of children in five categories, ranging from a normal group, which showed no trace of hookworm infection, to the most heavily infected group (from 90, p. 319).

Intensity of Infection	*Number of Cases*	*Average IQ*
Normal	17	90.2
Very light (1–25)*	40	88.3
Light (26–100)	27	86.4
Moderate (101–500)	23	84.1
Heavy (501–2000)	10	76.3

* Estimated number of hookworms.

Although the differences in averages are appreciable if extreme groups are compared, the overlapping of all groups is large. When individual scores rather than group averages are considered, a correlation of .30 [14] is obtained between IQ and degree of hookworm infestation.

This correlation, although not high, indicates a somewhat closer degree of relationship than has been found between mental level and any of the other physiological conditions so far discussed. The analysis of results obtained in investigations on hookworm suggests the operation of a factor which is probably present, although to a lesser extent, in all studies on the relationship between psychological and physical characteristics. The individuals of inferior physical condition in general tend to come from a *poorer socio-economic level*, their environment is deficient in opportunity for intellectual development as well as in sanitary conditions, facilities for medical attention, proper food and home care, and so forth. This is particularly well illustrated by hookworm, a condition which is relatively common among individuals of low social status and which flourishes in very poor and backward

[14] Computed by Paterson (65, p. 196), from the data of Smillie and Spencer.

rural districts. The environmental background may be the common underlying factor which leads both to the physical and to the psychological conditions. This could in itself account for what little relationship is found between physical condition and IQ.

Glandular Conditions.[15] It is well known that marked overactivity or underactivity of any of the endocrine glands may have a pronounced effect upon behavior. Within the range of normal variation, however, no significant relationship between glandular functioning and intellectual or emotional characteristics has been conclusively demonstrated (cf. 86). Among the most readily obtained indices of glandular activity is the familiar *basal metabolic rate* (BMR). This is a measure of the rate at which the body uses oxygen, which in turn depends upon the degree of activity of the thyroid gland. An abnormally low BMR can be raised by the administration of thyroid extract. Extreme underactivity of the thyroid results in cretinism, a condition characterized by feeblemindedness as well as by a number of clearly recognizable physical symptoms. The milder variations in BMR among normal adults or adolescents,[16] on the other hand, have consistently shown negligible or zero correlations with intelligence test scores in a number of investigations (cf. 86, p. 604). There is some evidence suggesting a significant relationship between BMR and scholastic achievement among college students, a relationship which may be attributable to individual differences in general energy level (58).

A possible association between mild glandular abnormalities and personality disorders is suggested by the relatively large incidence of glandular disorders among "problem children." In one survey (54) of 1000 children who were classified as behavior problems, 20% showed some glandular defect. In 10%, the glandular condition seemed to be a causal factor in the behavior disorder. That the relationship may not be so direct as these data imply is suggested by the variety of behavior disorders which are associated with the same type of glandular disorder. Conversely, the same kind of behavior disorder is found in children with entirely different glandular defects. As is true of many physical conditions, the relationship with behavior is general and not specific. A plausible hypothesis to account for the observed

[15] For a summary of the data pertaining to the effects of endocrine secretions upon behavior in man and animals, cf. Beach (9).

[16] For data on BMR in children, see the section on Developmental Relationships in a later part of the present chapter.

association between glandular disorders and behavior problems is based upon *indirect social effects* of the abnormal physical condition. If the glandular defect handicaps the child or renders him in any way different from other children, the behavior problem may simply represent the child's reaction to this abnormal situation.

"The Internal Environment." The physical and chemical condition of the blood, which constitutes the internal environment of the organism, is of prime importance in the normal functioning of the individual and in the maintenance of life itself. A large number of investigations have demonstrated pronounced behavior symptoms following changes in such conditions as the temperature, oxygen content, sugar content, or acid-base balance of the blood (4, 16, 86). One illustration is to be found in the well-known effects of oxygen deprivation—as in high altitudes—which include conspicuous alterations of sensory, motor, intellectual, and emotional responses (59, 60). There is evidence that some of the blood conditions which produce temporary disturbances of cerebral functioning may lead to irreversible changes in the brain cells and thus effect permanent behavior modifications in the individual. Especially significant are agents of this sort operating in early childhood or during prenatal life. Severe anoxia (oxygen lack) at birth, for example, may produce brain damage leading to motor, intellectual, or emotional disorders throughout life (cf. 86). Although some of these blood conditions may themselves be genetically determined, others probably depend upon characteristics of the individual's previous environment. These investigations thus suggest another way whereby environmental factors may influence the individual's subsequent behavior development.

A somewhat different question is whether individual differences in blood chemistry among normal adults are in any way related to behavior differences. It should be noted in this connection that the body has a number of regulatory mechanisms which preserve the stability of the internal environment within very narrow limits. The maintenance of this relatively stable state has been termed *"homeostasis."* One of the important regulatory mechanisms is provided by the action of various endocrine glands, which counteract chemical deficits or excesses in the blood composition. Owing to such internal safeguards, the composition of the blood does not vary widely among individuals or within the same individual *under ordinary conditions.* Despite this fact, hypotheses regarding the relationship between indi-

vidual differences in blood composition and intellectual or personality traits have been plentiful. The study of the behavior correlates of blood chemistry is today an active field of research, but so far the data have been contradictory and disappointing. One of the most widely discussed of these possible relationships is that between emotional stability and homeostasis. There is some evidence (25) which suggests that the more neurotic individuals tend to exhibit greater daily fluctuations in blood composition than the better-adjusted individuals.[17] At best, however, the results on blood chemistry and behavior provide interesting leads for future research.

A number of recent investigations have been concerned with "autonomic balance," by which is meant the interaction between the sympathetic and parasympathetic branches of the autonomic nervous system. Although this research is still in an exploratory stage, there is some evidence suggesting a possible relationship between physiological indices of autonomic balance and emotional and social characteristics in children (99).

Psychosomatic Disorders. A number of conditions such as asthma, skin allergies, and gastric ulcers have in recent years attracted considerable attention because of their possible "psychosomatic" origins. This simply means that psychological factors may serve as contributing and in some cases even determining conditions in the development of these physical disorders. Many descriptions of the so-called ulcer-type personality have appeared, although most of these descriptions are based upon the general impressions of clinicians rather than upon controlled observations (cf. 79). So common is the belief that worry, tension, and excessive drive are associated with stomach ulcers that this condition has sometimes been described as "Wall Street stomach." Less widely known are the theories proposed regarding other psychosomatic disorders. Some observers have suggested, for example, that allergic children are more intelligent or more dominant in social relations than non-allergic children (cf. 86). Other studies, however, have failed to corroborate any of these claims regarding allergies (86).

Whether significant relationships between intellectual or personality characteristics and any of these "psychosomatic" conditions will be found when more and better investigations are conducted remains to

[17] A further discussion of homeostasis will be found in a later chapter on sex differences (Ch. 19).

be seen. Should any such association be established, however, its interpretation would still present difficulties. Does the physiological condition lead to the behavior manifestations, or vice versa? Are both the result of certain environmental factors, such as occupation or socio-economic level? Are the psychological effects in part an indirect consequence of the social handicap occasioned by the physical condition? Any one of these relationships could theoretically hold. It is probable that all are involved to some extent.

NUTRITIONAL FACTORS

The serious food shortages in many countries following World War II have given special impetus to the study of the effects of malnutrition, and have made this a topic of major social concern. The rapid growth of the young science of nutrition and the extensive research on vitamins have also served to focus attention upon the amount and nature of food intake by the body. Apart from well-established physical effects, are there psychological effects of diet? Certainly no shred of evidence has been presented to support any claims of the diet faddists. That fish is "brain food" and meat makes a person more irritable and aggressive are examples of old wives' tales and no more. More worthy of serious study is the possibility that general malnutrition may have a deleterious effect upon intelligence. All surveys conducted *within the normal range* of either nutritional status or intelligence have shown only a slight positive correlation between these two variables (21, 38, 86, 87). Even this correlation tends to disappear when comparisons are made within a relatively homogeneous social group. The influence of socio-economic level upon these correlations is probably similar to that discussed earlier in connection with hookworm: the brighter children tend to come from better homes, which also provide more adequate diet.

Studies in which undernourished school children were given special diets for a period of several months and brought to a normal physical condition have yielded inconsistent results with respect to intellectual improvement. It is likely that when positive results are reported in these studies an uncontrolled motivational factor may have operated. For example, in one investigation (81), 50 underprivileged children were separated into two equated groups, one of which was served a special daily breakfast at school throughout the experimental period,

while the other was not. The performance of the experimental group improved more than that of the control group in both school work and standardized tests, but the difference diminished gradually after the breakfasts were discontinued. In this situation, the motivational effect of the special attention shown the experimental group could account for the entire difference in performance. The results are therefore inconclusive with respect to nutrition.

There is some evidence that nutritional status may be more closely related to intellectual performance among *mental defectives* than among persons of higher intellectual level. One investigator (45, 74) tested 41 feebleminded children, aged 2 to 7 years, who were undernourished at the time of the first test and well nourished at the time of the second test. The control group consisted of 41 uniformly well-nourished children, matched as closely as possible with the experimental group in chronological age, IQ, and interval between the two tests. Following their improved nutritional status, the experimental group gained an average of approximately 10 IQ points, while the control group showed no change during the same period. It required from 18 to 24 months of the dietary regimen to bring about the improvement in IQ. The age of the subject also affected the results, the greatest gains being made by children under 5.

It is also likely that when the diet is close to the *subsistence margin* a closer relationship may exist between nutritional level and intelligence (cf. 39). Among the most common effects of severe malnutrition are fatiguability, lack of energy, and lassitude. If prolonged, these conditions would themselves interfere with learning and thus retard intellectual development, even if no other effects of malnutrition on behavior are to be found. Similarly, malnutrition *in combination with other poor health conditions* may constitute a sufficiently serious handicap to interfere with normal behavior development in young children. In one investigation, intelligence tests were administered to children undergoing outpatient treatment for various disorders either in a clinic or in a private physician's office (45). Within this group, 50 who were classified as undernourished at the time of the first examination showed a significant rise in IQ after their nutritional status was brought up to normal. A control group of 50 well-nourished children who were also outpatients at the same centers showed no significant IQ change over a similar period of time. Since the subjects in this investigation were all patients, it is likely

that many of the malnourished children were either severely undernourished or were suffering from other health difficulties which made their total physical handicap more serious.

Nutrition research has demonstrated that the *qualitative aspects* of diet are even more important than the quantitative. Animal experiments, as well as clinical observations on humans, have furnished ample evidence that serious physical disorders may result from the lack of one or more essential vitamins from the diet. A lively area of current research is concerned with the psychological effects of vitamin deficiencies. Because of the known physiological effects of vitamins of the B-group upon the nervous system, most of the behavior studies have concentrated on this group. There seems to be good evidence that a deficiency in B-vitamins reduces physical strength and endurance (7, 13, 86). Clinical reports (cf. 86) on patients with vitamin B deficiency have consistently mentioned irritability, moodiness, and lack of cooperation. In cases of more severe deficiency, apathy, depression, and emotional instability are observed. Relatively few well-controlled experimental studies on the effect of vitamin deficiencies upon human behavior are available, and most of the investigations have dealt with too few cases to be conclusive. In general, these studies show no diminution of intellectual functions, but only motor and personality changes (7, 27, 86). Nor has the administration of excess vitamin B to normal individuals shown any consistent effects on behavior.

On the other hand, there is some evidence to suggest that the continued administration of *thiamin* [18] to children whose diet has been somewhat deficient in vitamins may lead to significant improvements in certain behavior functions. A well-controlled experiment (28) was conducted on matched pairs of orphanage children whose normal diet was relatively low in vitamin content. One member of each pair received regular thiamin pills, while the other received a placebo, or "bread pill," as a control. The procedure was such that neither the children nor any member of the orphanage staff knew which were the experimental and which the control children. Follow-ups over a two-year period showed a significant difference in favor of the thiamin-fed group in such tests as visual acuity, rote memory, and code-learning. The nature of the tests suggests that the advantage of

[18] One of the B-complex vitamins.

the thiamin-fed group may have resulted largely from greater alertness and better ability to concentrate.

A few investigations have been concerned with the effects of *glutamic acid*[10] upon psychological functioning. One study (104) reported a significant increase in IQ among 44 mentally retarded children following the administration of glutamic acid for six months. In another, 60 mental defectives were tested before glutamic acid therapy, as well as four and eight months following the beginning of treatment (23). As a control measure, part of the group received glutamic acid and part received a placebo at the start, the procedure being reversed for some of the subjects after four months. Preliminary results indicated a small but significant rise in IQ following the glutamic acid therapy, while no improvement followed the administration of the placebo. The permanence of the improvement noted in such investigations cannot, of course, be determined without more extensive follow-ups. Moreover, it is possible that the observed intellectual improvement may have resulted largely from an increase in alertness following the glutamic acid therapy

One of the few intensive and well-controlled experimental studies of the effects of nutrition upon human behavior is that conducted at the Laboratory of Physiological Hygiene of the University of Minnesota (11, 12, 26). Thirty-four men between the ages of 21 and 33, who volunteered for the experiment, were kept for six months on a semi-starvation diet described as characteristic of European famine conditions. As a standard, each subject's normal performance during a three-month period of adequate diet was recorded. The daily intake of calories averaged 3150 during the preliminary normal period and 1755 during the second, or experimental, period. The average weight loss of the subjects during the semi-starvation period amounted to about 25%.

The clearest behavior change during the experimental period was a decline in strength and endurance in motor tasks, and a less marked but significant loss in motor speed and coordination. No sensory effects were noted except a rise in auditory acuity and an increased sensitivity to cold. On a series of tests of intellectual functions, no change in either speed or level of performance was observed; nor was learning affected. In contrast to this lack of impairment as deter-

[10] Glutamic acid is one of the essential amino acids derived from proteins.

mined by objective tests, the subjects believed that they had deteriorated sharply. Self-ratings in alertness, concentration, comprehension, and judgment dropped markedly in the course of the experimental period. These differences are probably related to the personality changes, which were conspicuous. Personality tests (Minnesota Multiphasic and Guilford-Martin Inventory) showed a statistically significant increase in depression, hysteria, hypochondria, nervous symtoms, feelings of inadequacy and inferiority, and introversion; a decrease in general activity and in social leadership was likewise indicated. Constriction of interests and obsessive preoccupation with thoughts of food were very apparent.

The third stage in the experiment consisted of a 12-week controlled nutritional rehabilitation period, in which the caloric intake was increased by different amounts in different sub-groups. The experimental design also included groupings in which the diet was supplemented with protein or with vitamins. The weight rises during this period varied with the amount of caloric intake, but were not significantly related to vitamin or protein supplementation. The effects on motor, sensory, intellectual, and personality functions paralleled, in reverse, the previous changes during semi-starvation. The improvement was large in motor functions and in personality characteristics, but no significant change was found in intellectual functions. Vitamin and protein supplementation produced no significant differences in behavior recovery, although the amount of caloric intake did. This experiment indicates that the behavior changes induced by semi-starvation are reversible and remediable. It should be remembered, however, that such a finding applies to a six-month period of inadequate nutrition in adults. What would occur in a child, or following a longer privation period, we cannot infer.

DEVELOPMENTAL RELATIONSHIPS

In an earlier chapter (Ch. 5), we discussed the structural changes which parallel changes in behavior in the growing organism. The study of such changes represents a developmental rather than a statistical approach to behavior differences, and has been limited principally to prenatal or early postnatal stages of development. The studies which have been considered in the present chapter, on the other hand, have been concerned with correlations between struc-

tural and behavior differences in relatively mature organisms. These distinctions and classifications become less sharp as the intermediate areas which bridge the gap between isolated fields of research are developed. Correlational studies of physical and psychological characteristics among *children* are an example of such a marginal area of investigation, since both individual differences and developmental differences within the individual contribute to these relationships.

In the case of traits which show appreciable age changes, any correlations found among children should be considered apart from similar correlations obtained in adult groups. A relationship present in the growing organism may disappear when maturity is reached, since it may have resulted simply from developmental influences. It is obvious, of course, that a 10-year-old will excel a 5-year-old in both arithmetic and height. Thus if 10- and 5-year-olds are included within the same group, an artificial correlation will be obtained between arithmetic and height. Such a "spurious" correlation is usually eliminated through the use of relative measures (e.g., IQ) or through comparisons within a single age group. But these procedures do not rule out the entire contribution of developmental differences, since children of the same chronological age may vary widely in the degree of physical development which they have attained.

There is no consistent relationship between developmental rate and adult status in physical characteristics. The data on age of onset of puberty furnish a good illustration of this point. Individuals who reach sexual maturity earlier are generally accelerated in physical development from early childhood (89). The age of onset of puberty is thus one manifestation of the individual's general rate of physical development. During childhood, the earlier-maturing individuals will be taller, heavier, and farther advanced in most physical characteristics than those who reach puberty later. But in adulthood, those who reached puberty earlier are *not* taller or heavier. In fact, a slight tendency has been found for earlier-maturing girls to be somewhat shorter during the late teens (89, 96). The tallest child in a group will not necessarily be the tallest twenty years later. The physical status of a child depends in part upon certain absolute factors which make some individuals, for example, taller than others throughout life, and in part upon individual differences in developmental rate.

In the light of these considerations, it is perhaps not surprising to find that correlations between anatomical or physiological charac-

teristics and intelligence tend to run higher among children than among adults. These correlations are still quite low, rarely exceeding .30, but they are often high enough to indicate a statistically significant relationship.[20] In an extensive study by Abernethy (1), for example, positive correlations were found between various anatomical measures and intelligence at all ages from 8 to 17, but the correlations tended to be lower in the groups which were approaching maturity. In a comparable adult group included in this study, the correlations were virtually zero. With young children, ranging in age from 3 months to 8 years, Bayley (5) found average correlations of .16 and .15 between the height and intelligence of boys and girls, respectively. There is also some evidence that "skeletal age," as determined by X-ray photographs of bone structure, is significantly correlated with intelligence in children, and that the correlation diminishes with age (18).

Some startlingly high correlations have been reported by Hinton (33) between basal metabolic rate and IQ in a group of 200 children ranging in age from 6 to 15. These correlations were close to .80 for the 6- to 9-year-old groups; from age 10 on they dropped fairly consistently, reaching a value of about .50 among the 15-year-olds. It will be recalled that investigations on adolescents and adults showed virtually zero correlations between BMR and intelligence. If Hinton's results are confirmed by other studies, they may provide an interesting illustration of age changes in the relationship between bodily conditions and behavior. It may be noted that the BMR tends to be higher during periods of rapid growth. If BMR is shown to be significantly related to intellectual level in childhood, this may help to explain many of the other correlations.

It has been argued that, in both their physical and psychological development, some individuals may progress at a more rapid rate than others throughout their period of growth. According to this hypothesis, it is these differences in *developmental rate* which may account for the slight positive correlations found between intelligence and certain bodily characteristics among children. It should be remembered, however, that growth does not occur at a uniform or regular rate within the individual, but exhibits many irregular spurts and lags. These temporary fluctuations in rate of growth are quite

[20] Cf. 18, 36, 62, 67, 88.

specific, and no parallelism has been discovered between psychological and physical fluctuations within individual growth curves. Thus the monthly or annual *increments* in structural and in intellectual status are generally uncorrelated (1, 18). This suggests that whatever relationship exists between bodily and behavorial development is probably an indirect one. For example, the child who is physically accelerated is likely to learn to walk—and possibly talk—earlier, thereby expanding his environmental contacts in advance of the slower-maturing individual. This could account for a slight difference in intellectual development in favor of the earlier-maturing child. On the other hand, the temporary ups and downs in physical and psychological development seem to result from a multitude of unrelated factors, and offer no support to the theory of a "common underlying growth tendency."

The effect of puberty upon behavior development has itself been widely discussed. Contrary to popular belief, there is no evidence that intellectual development is either consistently accelerated or hindered by the onset of sexual maturity (1, 18). Nor is there any relationship between age of sexual maturity and either intellectual or personality characteristics in adulthood, when racial and cultural differences are held constant (1, 96). The onset of puberty does, in general, usher in changes in attitudes, interests, and emotional reactions. In one survey (97), significant differences were found between the personality test responses of pre-pubertal and post-pubertal girls of the same chronological age and comparable socio-economic and cultural status. The important role of social factors in bringing about these personality changes cannot however, be overlooked.

SENSORY HANDICAPS

Sensory limitations have a much more direct bearing upon behavior than most other kinds of physical deficiency, since they cut off environmental stimulation. The individual so afflicted is psychologically "isolated" from cultural contacts in the same sense as the wolf children of Midnapore or Kaspar Hauser (cf. Ch. 6). We should therefore expect a fairly pronounced behavioral deficiency to be associated with sensory defects. For man, visual and auditory defects are obviously the most serious sensory handicaps. Since human culture is built to such a large extent upon a foundation of language—a

language acquired principally through the eye and the ear—deficiencies in these sensory areas are of basic significance.

Visual Handicaps. Any over-all estimate of the average intellectual status of the visually handicapped is extremely difficult and probably meaningless (cf. 29). One reason is that intelligence tests devised for the blind are *not usually comparable* to those for sighted children because of the omission and substitution of tests, changes in administration, and other alterations necessitated by the visual handicap. The most nearly comparable test is Hayes' recent adaptation of Forms L and M of the 1937 Stanford-Binet for blind children (31). For blind adolescents and adults, the Wechsler-Bellevue can be used with only minor modifications (30). These tests have not been in use very long, however, and most of the large-scale comparisons between blind and sighted are based on earlier, less comparable tests.

Another point to consider in such comparisons is the *cause of blindness*. In some cases, blindness results from pathological conditions which also lead to neurological deterioration. If the IQ's of such individuals are included in the total estimate, they simply confuse the picture. The intellectual achievement of blind children also depends upon the *amount and nature of special education* which they have received. Such training tends to compensate for the visual handicap by providing the necessary contacts with the social environment through other sensory channels. With the marked progress in methods of instruction for the blind, it is likely that the average IQ of children in blind schools today is higher than it was twenty years ago—and that twenty years hence it will be still higher.

The *age of onset of blindness* is likewise related to the amount of intellectual handicap, although the relationship is not a simple one. On the one hand, the later the loss of vision occurs, the more opportunity the individual will have had for normal educational experiences. On the other hand, such an individual will have had less time to adjust to the blindness, and may encounter more interference in the acquisition of the new reaction systems required by the loss of vision. These two opposing influences probably account for the lack of correspondence generally found between age of onset of blindness and intelligence test performance or educational achievement (29). Another factor contributing to the intellectual development of the blind child is his *emotional response* to the handicap. The attitudes of his family and

associates, the general nature of the home milieu, and many other attendant circumstances will determine how effectively the individual adjusts to the handicap and will indirectly affect his educational and intellectual progress (cf. 91).

A final and very important consideration in connection with the general intellectual level of the visually handicapped is the *extent of the defect*. Like all psychological characteristics, vision tends to follow a normal distribution in the general population. Between the large "normal" group and the totally blind, one finds innumerable degrees of handicap along a virtually continuous scale. Sharply distinguished categories are just as out of place here as in other aspects of individual differences. For practical convenience, a threefold classification is now generally employed, including correctable defects, partially seeing, and blind. The per cent of children falling into each of these categories, as estimated by the White House Conference on Child Health and Protection (101), is as follows:

Correctable defects	19.75
Partially seeing	0.20
Blind	0.05

Correctable visual defects, when actually corrected by the use of glasses, have no effect upon intellectual development. If the child wears glasses from the time when the defect becomes appreciable, no interference with normal environmental contact results. When the defect is not compensated by means of lenses, however, the child's school work, and indirectly his intellectual development, usually suffer. Inattention, lack of interest in school, loss of self-confidence, and inferior performance may result from unsuspected visual deficiencies.

The term "*partially seeing*" is applied to children whose visual deficiency is so serious [21] as to necessitate special instructional techniques in sight-saving classes, where classroom procedures are adapted to a limited use of vision. Surveys [22] (3, 101) in such sight-saving classes have shown an average IQ of about 90. The distribution is quite skewed, with a marked piling up of cases at the lower IQ levels. About 50% of the children have IQ's below 90; and of these, from 6% to 10% are below 70. Less than 10%, on the other hand, have

[21] Roughly, the limits are between 20/70 and 20/200 vision, although other factors are also taken into account.

[22] In one of these surveys (3), an adaptation of the Stanford Binet was employed in which certain tests were reproduced in magnified form and with heavier and darker lines; in the other (101), the test is not specified.

IQ's of 110 or higher. At the time of their admission to sight-saving classes, many of these children show personality disorders associated with their visual handicap (3). Introversion, daydreaming, feelings of inferiority, and tension have been most frequently reported.

The *blind* have been defined as those who cannot be educated through visual means. Within this group, the individuals with a little vision are sometimes the most retarded (29). Because they are less highly motivated to learn to depend upon touch, such individuals tend to dissipate their efforts. Certain selective factors operating in admission to schools for the blind also tend to produce a difference in the same direction. Thus a bright child with a marginal amount of vision is more likely to succeed in a sight-saving class, while a dull child with the same amount of vision may fail and be sent to a school for the blind. Several surveys conducted in schools for the blind have shown an average retardation of from two to three years in school progress, but only a slight retardation in intelligence test performance. As nearly as can be estimated, the average IQ of blind children is slightly above 90, and the per cent of IQ's in the subnormal levels is about twice as large as that in the superior levels (29, 71). On the whole, the intellectual handicap seems to be no worse than that of the partially sighted, and there is some evidence that it is slightly less.

In such tasks as learning a maze, the blind do somewhat better than blindfolded normal subjects (29, 71), probably because of the greater familiarity of the blind with the use of non-visual cues. There is no evidence, however, for the popular belief that the blind have a finer discrimination than the sighted in other senses, such as hearing or touch (80). The remarkable feats often accomplished by blind persons through the use of other senses stem from a more efficient use of sensory cues rather than from a superiority of the senses themselves. Through prolonged training, an individual may acquire the ability to respond to very slight cues which are ordinarily ignored. Such seems to be the case among the blind. The so-called obstacle sense of the blind, which enables them to perceive obstacles in their path, has been shown to be based primarily upon learned responses to auditory cues (102).

In personality development, the adjustment made to the visual handicap varies widely with the individual. The range of personality characteristics is fully as wide among the blind as among the sighted. There is some evidence (10, 61) that the number of neurotic symp-

toms among blind adolescents is significantly greater than among comparable sighted groups, but the difference is smaller than might be expected in view of their handicap. Emotional maladjustment tends to be less among the totally blind than among the partially sighted (61). Moreover, it seems fairly clear that it is not the defect itself, but the social treatment, which is at the basis of the insecurity and other emotional difficulties of the blind (52, 53).

Auditory Handicaps. Contrary to popular belief, hearing deficiencies constitute a more serious handicap to intellectual development than do visual defects. Deafness interferes more than blindness with language development and with normal social contacts. In estimating the intellectual handicap occasioned by hearing deficiencies, the same difficulties are encountered as in the testing of the visually handicapped. A working classification of auditory deficiencies has been devised which closely parallels the threefold classification of visual handicaps discussed above. It has been estimated that about 14% of school children have *defective hearing*, a category referring to milder hearing disabilities (3). Such handicaps more often escape notice than the visual, and the child's behavior is mistaken for carelessness, indifference, rudeness, or dullness. Among the effects of such handicaps upon the child are poor scholarship, speech defects, loss of interest, social aloofness, and suspicion (3).

Those classified as *hard-of-hearing* have a more conspicuous defect, but are nevertheless able to use hearing in acquiring an understanding of spoken language.[23] In general intelligence, language development, and educational progress, they are intermediate between those with minor hearing deficiencies and the deaf (3, 71, 72, 95). There is some evidence to suggest that on such tests as the Pintner Non-Language Scale, hard-of-hearing school children are not inferior to their normal-hearing classmates (20). When hard-of-hearing children are matched with normal-hearing classmates in non-language intelligence test score, however, the hard-of-hearing do more poorly than the normal-hearing in tests of scholastic achievement (20).

Among the *deaf* are included those individuals whose hearing deficiency is so serious as to prevent the acquisition of language in the ordinary environment. Formerly known as "deaf-mutes," such individuals provide a vivid demonstration of the influence of environ-

[23] Either because the handicap is less severe than that of the deaf or because the loss of hearing occurred after the acquisition of language.

mental stimulation upon the development of an important behavior function. Never having heard the human voice, the "deaf-mute" is unable to speak, although his vocal organs are perfectly normal. The presence of human vocal organs does not in itself lead to the development of human speech, any more than any other structure insures the appearance of the function ordinarily associated with it. Vocal organs of a certain type are a necessary but not a sufficient condition for the acquisition of speech. That the deficiency of the "deaf-mute" is a stimulational one is shown by the fact that, with modern teaching methods based upon the use of other sensory cues, such individuals can be taught to speak normally. The remarkable results achieved with certain persons who were both blind and deaf point up still further the importance of training in behavioral development. The most famous examples are Helen Keller and Laura Bridgman, who achieved considerable eminence in their busy careers despite this dual handicap.

A number of extensive test surveys have been conducted in schools for the deaf (cf. 71). Educationally, such groups are as much as four or five years retarded. On the usual verbal-type intelligence test, the deaf experience considerable difficulty because of their deficient mastery of language and linguistic concepts. So great is this handicap, that verbal tests are generally considered inapplicable to deaf children, even though such tests may involve no spoken language. The problem of testing the deaf was, in fact, one of the principal reasons which led to the construction of the early non-language and performance scales. In one of the most comprehensive surveys of deaf children (cf. 71, p. 118), the Pintner Non-Language Test was given to 4432 children, 12 years of age or older, in 41 schools for the deaf. The average MA and IQ of the children in each year group from 12 to 15 were as follows:

Chronological Age Level	*Average MA*	*Average IQ*
12	10–9	86
13	11–2	84
14	11–8	83
15	12–1	82

On performance tests, the mean IQ of deaf children ranges from slightly below 90 to slightly above 100, depending upon the nature of

the tests.[24] It is likely that this difference is largely a matter of the degree to which language concepts aid in the performance of the test. Since language serves an important function in so much of our thinking, the linguistic retardation occasioned by deafness handicaps the individual in a fairly broad area of intellectual activity. No relationship has been found between the age of onset of deafness and scores on non-language or performance tests. Educational achievement, on the other hand, is clearly better when the loss of hearing occurs at about 4 years or older, after the normal acquisition of language (71).

In personality development, the deaf and hard-of-hearing show many of the same reactions noted among the visually handicapped. They tend as a group to be somewhat more emotionally unstable, introverted, shy, and insecure than normal-hearing persons (14, 56, 69, 71). Deaf children also tend to be below the norms in social maturity, are less likely to be leaders, and present more behavior problems than other children (63, 71). In general, the more severe the hearing handicap, the greater the personality maladjustment (71), a relationship which did not hold in the case of the visually handicapped. It is interesting to note that deaf children who come from homes in which there are deaf adults tend to be better adjusted than those reared in homes in which all the adults have normal hearing (70). This suggests the dependence of the deaf child's emotional adjustment upon proper adult understanding of the child's problems during his formative years.

GENERAL EVALUATION

Our fundamental question in this chapter has been: how are individual differences in behavior related to individual differences in bodily conditions? Let us see what sort of an answer the data have provided. First, we must recognize that certain pathological conditions of the organism have characteristic physical as well as behavioral symptoms. But we cannot generalize from the association found in these abnormal cases to a possible connection within the normal range of variation. To take an obvious and extreme illustration, a person whose legs have been amputated at the knee is usually unable to dance. It does not follow, however, that length of leg is correlated with ability to

[24] Cf. 15, 46, 55, 63, 71, 93, 103.

dance and that those persons with longer legs will make the better dancers.

Aside from the relationships which have been demonstrated in pathological conditions, surprisingly little is known directly—although much has been inferred—regarding the operation of physiological factors in behavioral development. In the field of endocrinology, for example, much remains to be learned. Too often what has been offered as a stimulating hypothesis for further research has been interpreted by the layman as an established fact. The same may be said in regard to nerve physiology. The field abounds in speculation and the experts still disagree. At such a stage, it is definitely premature to venture a systematic analysis of behavior differences in terms of the nervous system.

Turning from the observation of pathological cases and speculations on the physiological mechanisms underlying behavior to data collected on normal groups, we still meet difficulties. Many of the investigations on this problem have been inadequately controlled. Through the misinterpretation of statistical techniques, slight general trends in groups have been erroneously attributed to individual cases. It will be recalled, for example, that the small differences in group averages, which were regarded as significant by many early workers, actually showed only a negligible relationship when the individual scores were correlated. The pronounced overlapping of groups was often ignored. Age differences were occasionally present within the groups, thus producing a spurious connection between certain physical characteristics and intellectual level. In many investigations showing a relationship between physical condition and intelligence, differences in *socio-economic level* may account for whatever positive correlation has been found. The individual who comes from a better home will have richer opportunities for intellectual development and at the same time will receive better physical care. He will be brought up under more sanitary conditions and will have less chance of contracting disease than the less fortunate child reared in a city slum or a poor rural district.

All in all, the available data furnish little evidence for a *bona fide* connection between behavior characteristics and physical conditions among normal persons. Many fields of research within this area, however, have barely been touched, and more information is obviously needed for a definitive answer. The *explanations* which have been proposed for such relationships fall into four major types. The theory

which prompted the investigations of Galton and other early workers is that of *"constitutional" inferiority or superiority.* In contrast to the popular notion of compensation, this theory maintains that "good things go together" and that the person who is superior in one respect tends to be superior in others. Any positive correlation between psychological and physical traits is attributed to a common "quality of the organism" which underlies all forms of development. Recent studies lend practically no support to this interpretation.

A second possible type of relationship may be described as *direct behavioral handicap.* This is best illustrated by pathological extremes, such as cerebral anoxia, the extremely small brain of the microcephalic idiot, or the underactive thyroid of the cretin. In such cases, the "minimum structural prerequisites" for normal behavior development are absent. It is becoming increasingly apparent, however, that among the large majority of individuals the direct control of behavior by structural factors is not very rigid. Beyond a certain essential minimum, further differences in structural characteristics are not necessarily accompanied by corresponding differences in behavior. To put it differently, the structural equipment of most individuals permits a very wide latitude in behavior development.

Another explanation is based upon the *indirect behavioral handicap* resulting from physical deficiencies. This handicap can take many forms. In the case of sensory deficiencies, there is a partial stimulational isolation of the individual. Malnutrition, poor health, and other general physiological conditions reduce endurance, increase fatiguability, affect muscular development, and generally lower the efficiency of work. These conditions, if sufficiently prolonged, may be expected to retard intellectual development to a certain extent. Physical defects or discomforts also serve as a powerful distraction and thus make it more difficult for the child to concentrate on his school work or other tasks. Finally, certain striking facial, cranial, or bodily characteristics which have acquired a specific significance through social stereotypes may affect the individual's subsequent intellectual and emotional development, because of the attitudes which they engender. The social consequences of poor health or of sensory handicaps exert a similar influence upon behavior.

A fourth and final type of relationship, which has come to the fore in recent years, is that implied by the term *"psychosomatic."* In this case, the psychological condition is logically regarded as the ante-

cedent, and the physiological disorder as its consequent. One of the clearest examples is to be found in the digestive, circulatory, and other internal changes occurring during emotional excitement. It is reasonable to suppose that continued stimulation of such physiological reactions may lead to a more lasting disruption of function, as illustrated by the development of gastric ulcers.

In conclusion, the evaluation of any investigation purporting to show a relationship between physical and psychological characteristics involves two questions. First, is the relationship genuine, or does it result from socio-economic or other uncontrolled conditions? Secondly, when a genuine relationship has been demonstrated, what is its specific nature, and how can the relationship be explained?

CHAPTER

13

The Quest for Constitutional Types

THE RELATIONSHIP between physical and psychological traits has also been considered from the point of view of *constitutional types*. In the effort to simplify the almost infinite observable variations among individuals, certain basic human types have been proposed. A specific individual can then be described as a more or less close approximation to one of a small number of types. Such constitutional types are offered as a characterization of the individual as a whole, in all his physical, intellectual, and emotional traits, and are not to be envisaged in terms of any isolated qualities of the organism. There is also a strong presumption of an innate or hereditary basis for the development of types. Thus a theory of constitutional types implies a certain degree of conformity among the various characteristics of the individual, these characteristics being ultimately attributed to an underlying innate tendency.[1]

Type theories have been eagerly received by the general public as a short-cut to the understanding of human nature. The layman is impatient with the slow, meticulous methods of science. This is particularly true in psychology, because of the more intimate and immediate bearing which this science has upon man's everyday life. The terminology of type theories has become such an integral part of our language that it is almost impossible for us to speak about people without inadvertently lapsing into hypothetical categories. The popular tend-

[1] The concept of types has also been employed in the description of specific functions, as in Galton's classification of individuals in regard to their predominant field of imagery, i.e., visual, auditory, olfactory, etc (12). Such types, however, do not characterize the personality as a whole and are not to be confused with the constitutional types under consideration.

ency to make sharp distinctions, coupled with the hope that character and mentality can be "read" from physical signs, has helped to keep "types" alive. Among psychologists, there have been recurrent revivals of interest in typology. As new theories appear, they are followed by a flurry of hopeful research. In the sections which follow, we shall consider some of the best-known type theories, inquire into their psychological implications, and examine some representative data collected to support or test their claims.

TYPE THEORIES THROUGH THE AGES

The first clearly formulated attempt to classify individuals into basic types was probably that of the Greek physician Hippocrates in the fifth century B.C. Hippocrates proposed a twofold division into *habitus apoplecticus* and *habitus phthisicus*. The former corresponds to a thick-set, heavy body build, susceptible to apoplexy and similar physical disorders; the latter is characterized by a long, slender body and susceptibility to respiratory diseases such as tuberculosis. Because of the predominantly medical interest of its exponent, this classification was based primarily upon relative susceptibility to different kinds of physical ailments. Such an approach has, however, persisted to the present, many current type theories taking susceptibility to various physical or mental disorders as their starting point.

The second-century Greek physician Galen, frequently called the father of modern medicine, is responsible for the well-known classification of "temperaments" into the sanguine, the choleric, the phlegmatic, and the melancholic. These terms have achieved great popularity as descriptive figures of speech, and one wonders how often they are still being taken literally. The theories of both Hippocrates and Galen were founded upon a biochemical approach to personality. Thus Hippocrates attributed the development of his two types to the relative proportion of "fire" and "water" elements in the individual's make-up. Galen ascribed his four temperaments to the excess of one or another of four "humors" or body fluids.

In more modern times, many variations of type concepts have appeared in literature, art, philosophy, medicine, anthropology, and any other field in which man is the central figure. Every school child is familiar with the quotation from Act I of Shakespeare's *Julius Caesar*:

Let me have men about me that are fat;
Sleek-headed men and such as sleep o' nights.
Yond Cassius has a lean and hungry look;
He thinks too much: such men are dangerous.

The nineteenth and early twentieth centuries have been described as the golden age of typologies (38). The English anthropologist Walker,[2] in 1852, wrote of "nutritive beauty," "locomotive beauty," and "mental beauty." In the following year, Carus,[2] a German zoologist, described three bodily types: the phlegmatic, in which the region of the digestive organs is prominent; the athletic, with strongly developed bones and muscles; and the asthenic, with narrow chest, a long body, and poorly developed skeleton and musculature. In France, several type theories were proposed, chief among which was that formulated by Rostan (35) in 1828 and later adopted by Sigaud[3] and his students. This classification recognized four types: digestive, muscular, respiratory, and cerebral. Manouvrier[3] suggested a division into makroskele and brachyskele, or narrow skeleton and broad skeleton. MacAuliffe[3] offered the *type plat* (flat) and the *type rond* (round).

In Italy, Viola (cf. 30) formulated a theory which became familiar to psychologists through the researches of Naccarati (30, 31) and others in America. Viola's types include the macrosplanchnic, the normosplanchnic, and the microsplanchnic. The macrosplanchnic possesses a large trunk which is overdeveloped in comparison with the length of the limbs; the horizontal dimensions are relatively large, the vertical relatively small. The microsplanchnic, on the other hand, has a small trunk and long limbs, the vertical dimensions being relatively in excess of the horizontal. Between these two extremes is the normosplanchnic, which exhibits a proportionate and harmonious development of trunk and limbs. Viola suggested a series of body measurements to be employed in classifying individuals into these types. Naccarati (30) later substituted a single numerical expression of body build, the *morphologic index*, computed as follows:

$$\text{MI} = \frac{\text{length of one arm} + \text{length of one leg}}{\text{volume of trunk}}.$$

The trunk volume in this formula is determined by a series of rather elaborate measurements.

[2] Cf. Wertheimer and Hesketh (46).
[3] Cf. Wertheimer and Hesketh (46).

According to Viola's theory, the macrosplanchnic represents an overdevelopment of the nutritional or "vegetative system" contained within the trunk. The microsplanchnic, on the other hand, is characterized by an overdevelopment of the "animal system," consisting of the musculature, nervous system, and skeleton. The two types were believed to differ in intellectual and emotional characteristics as a result of the relative activity of the vegetative and animal systems, which were regarded as independent and even antagonistic in their action. In his elaboration of Viola's theory, Naccarati (30) suggested that the microsplanchnic corresponds to a hyperthyroid condition and should therefore be expected to manifest the various characteristics associated with overactivity of this gland.

Pende (cf. 46) subsequently proposed a distinction between hypervegetative and hypovegetative biotypes, a classification which, as the terms imply, has much in common with Viola's theory. A definite endocrine basis was offered for this distinction.

In America, Davenport (10) classified individuals into the fleshy, the medium, and the slender biotypes. Stockard (42) distinguished between the linear and the lateral types, which he related to the activity of the thyroid. The linear type was described as active, energetic, and nervous, but emotionally controlled; such individuals grow rapidly and reach puberty at a relatively early age. The lateral type is less active and grows at a slower rate. The linear type is also characterized by a dolichocephalic skull, the lateral by a brachycephalic one. Mention should also be made, from the psychological side, of the famous distinction proposed by William James (18) between "tender-minded" and "tough-minded" persons, a distinction which bears a certain resemblance to the introvert-extrovert classification to be discussed shortly.

Pavlov (32), the Russian physiologist of conditioned reaction fame, suggested a type classification in terms of the nervous system. On the basis of observations made in the course of his conditioning experiments on dogs, he proposed two predominant, opposed types, corresponding to extreme tendencies toward excitation or inhibition, respectively. Intermediate, less pronounced types were also described. Pavlov called attention to the resemblance between the classification so obtained and the classical division into sanguine, melancholic, phlegmatic, and choleric temperaments. He suggested that, "Until a rigid scientific classification is fully established for all the various

types of central nervous system . . . we may be permitted to make use of the ancient classification of the so-called temperaments" (32, p. 286).

Within the present century, type psychology has flourished most vigorously in Germany. Several variations and ramifications of type theory have been formulated by contemporary German psychologists. Jaensch (cf. 22, 24) proposed a classification of constitutional types on the basis of *eidetic imagery*. The eidetic image is a peculiarly vivid and detailed memory image [4] which is experienced by some individuals. Eidetic imagery has been found to be most common in late childhood and to disappear gradually during adolescence, although it has also been discovered among some adults. The eidetic image may be a photographic replica of the original object, or it may differ from the latter in certain characteristic ways. Jaensch recognized two types of eidetic individuals. In the first type, the image can be called up, banished, and altered voluntarily. The eidetic image in such cases may be nothing more than a "visualized idea" and it is accepted as natural and normal by the individual. In the second type, the image usually arises spontaneously and may persevere in spite of efforts to banish it; voluntary alterations in the qualities of the image are often impossible. Such images do not come up very frequently, and are often regarded as unpleasant and even uncanny by the subject.

Jaensch considered these two eidetic types to be distinct constitutional types, differing in many bodily and psychological traits and characterized by basically dissimilar "psychophysical reaction systems." The eidetic characteristics were simply taken as convenient starting points in the classification. The first of the two types described above was designated the B-type, because of its alleged resemblance to the Basedow syndrome,[5] and the second, the T-type, owing to the similarity of some of its manifestations to the condition of tetany.[6]

Jung's *introvert* and *extrovert* types are well known (19). Jung maintained that in the extrovert the "psychic energy" is turned outward to the objective environment; in the introvert, it is turned inward

[4] Eidetic images have usually been investigated in the visual field, although it has been claimed that they are equally common in other senses.

[5] A condition characterized by prominence of the eyeballs, enlargement of the thyroid gland, muscular tremors, rapid heart action, and more or less profound mental disturbance; believed to be caused by overactivity of the thyroid gland.

[6] A motor disorder, including muscular tremor, muscular spasms, and sometimes uncoordinated muscular contractions following upon an effort to make a voluntary movement; attributed to insufficient secretion of the parathyroid gland.

to a subjective world. The extrovert is predominantly oriented in all his actions, thoughts, interests, and feelings by the objects and people about him. His beliefs and opinions are guided by the mores of his group. The introvert, on the other hand, is governed by subjective factors; all his behavior has a subjective, inner reference. Jung regards these two types as fundamental biological contrasts. They denote for him basic attitudes which characterize all aspects of the individual's psychological make-up.

Jung's types have become more widely known, however, in terms of their emotional and social manifestations. Thus the introvert is usually thought of as an emotionally shut-in individual who shuns social contacts, prefers to work alone, and finds more pleasure in imaginative work than in a life of action. The extrovert suggests the "salesman" type, who meets people easily, is happiest in a social situation, and is friendly and interested in his fellow-beings. Jung regards introversion and extroversion as characterizations of *normal people*. In extreme forms, to be sure, they would predispose the individual to mental disorders which are opposite in their symptoms.[7] The fundamental distinction, however, is not made on the basis of these mental disorders. The susceptibility to one or the other form of insanity is considered simply another manifestation of the basic type.

Mention may also be made of Spranger's (cf. 41) description of six fundamental types of individuality, including the theoretical, economic, aesthetic, social, political, and religious. These "types" are regarded as meaning-tendencies or values in terms of which an individual's responses to his environment are to be understood. They are ideal types or schemata of understanding, rather than empirically observable types.

Kretschmer's type theory (25) has undoubtedly been one of the most influential in stimulating research. Physically, Kretschmer classifies individuals into four groups, the pyknic, athletic, leptosome, and dysplastic. The *pyknic* type of body build is short and thick-set with relatively large trunk and short legs, round chest, rounded shoulders, and short hands and feet. The *athletic* has a more proportionate development of trunk and limbs, well-developed bones and muscles, wide shoulders, and large hands and feet. The *leptosome* is generally

[7] This distinction was emphasized by McDougall, who wrote: ". . . persons of the extrovert temperament seem more liable, under strain, to disorder of the hysteric or dissociative type, those of introvert, or shut-in, temperament to disorder of the neurasthenic type" (28, p. 28).

characterized by small body volume in relation to height. He is tall and slender, with relatively narrow chest, long legs, elongated face, and long, narrow hands and feet. In the *dysplastic* category are placed all individuals who present an incompatible mixture of type characteristics in their physical development. Kretschmer suggested a wide variety of physical measures, to be used in conjunction with the clinical diagnosis of the experimenter, for differentiating between these bodily types.

The basic contention of Kretschmer's theory is that a relationship exists between the body types which he describes and two essentially opposed "temperaments," the *cycloid* and the *schizoid*. The cycloid individual manifests personality traits which in extreme cases would be classified under the cyclical, or manic-depressive, form of insanity. The schizoid tends toward schizophrenia, which is characterized by extreme introversion and lack of interest in one's surroundings. Kretschmer claims that the cycloid is usually pyknic, whereas the schizoid is leptosome or, less frequently, athletic. Although originally applied to different forms of mental disorders, this theory was subsequently extended to include normal individuals who manifest no personality disturbance. The terms "cyclothyme" and "schizothyme" were devised to denote these two normal biotypes. The former is described as social, friendly, lively, practical, and realistic; the latter as quiet and reserved, more solitary, timid, and shut-in. It will be noted that these descriptions correspond quite closely to Jung's extrovert and introvert types.

The latest revival of interest in typology followed the proposal of a somewhat different approach by Sheldon and his collaborators (39, 40) in this country during the early 1940's. This is not a type theory in one sense of the term, since individuals are regarded as falling along a continuous distribution in both bodily and psychological characteristics. What Sheldon argued for was a fundamental and probably innate relationship between body build and personality, neither of which need fall into distinct categories. We might say that this theory has retained the "constitutional" concept but dropped the "type" concept of the traditional typologies. A fuller discussion of Sheldon's theory, together with an analysis of the evidence for it, has been reserved for a later section of the present chapter.

Throughout the various type theories which have been described, we can detect a general dichotomy between two opposed constitutional

types. From the standpoint of physique, the distinction is one between the long narrow body, with relatively long limbs, and the short stocky build, with relatively large trunk and short limbs. In respect to personality, we are offered at the one extreme the expansive, sociable, easygoing, and practical man, and at the other the more taciturn, unsociable, intellectually independent, or idealistic type. Occasionally, more than two categories are given, but the additional types are usually found to be either intermediate degrees or modifications of the major ones.

In some theories, the structural classification is emphasized; in others the behavioral one is foremost. Many of the theories draw upon pathological conditions either for striking examples or for their basic concepts. Thus we frequently find susceptibility to a given class of physical or mental disorders as an outstanding characteristic of each type. In many cases, too, the various physical and personality types have been linked with race, and attempts have been made to attribute racial differences to the predominance of one or another constitutional type within each race.[8]

THE LOGIC OF CONSTITUTIONAL TYPES

Multimodal Distribution. Type theories have been most commonly criticized because of their attempts to classify individuals into sharply divided categories. Such a procedure would imply a multimodal distribution of traits. The introverts, for example, would be expected to cluster at one end of the scale, the extroverts at the other end, and the point of demarcation between them would be clearly apparent. Actual measurement, however, reveals a unimodal distribution of all traits, which closely resembles the bell-shaped normal curve (cf. Ch. 3). Moreover, it is often difficult to classify a given individual definitely into one type or the other. The typologists, when confronted with this difficulty, have frequently proposed intermediate or "mixed" types to bridge the gap between the extremes. Thus Jung suggested an ambivert type which manifests neither introvert nor extrovert tendencies to a predominant degree. Observation seems to show, however, that the ambivert category is the largest, and the decided introverts and extroverts are relatively rare. The curve, too, has no clear breaks,

[8] Cf., e.g., Weidenreich (45). For a further treatment of this application of typology, see Chapter 20.

but only a continuous gradation from the mean to the two extremes. As was indicated in Chapter 3, this general type of distribution has been found in practically all measurable traits of the individual, whether social, emotional, intellectual, or physical.

It is apparent, then, that in so far as type theories imply the classification of individuals into clear-cut classes in either physique or personality, they do not fit the facts. Such an assumption, however, is not necessarily inherent in all systems of human typology. It is more characteristic of the popular versions and adaptations of type theories than of the original concepts themselves. To be sure, type psychologists have often attempted to categorize individuals, but this was not an indispensable part of their theories; their concepts have occasionally been sufficiently modified to permit a normal distribution of traits.

It has been suggested, for example, that types may refer simply to original varieties, breeds, or "biotypes" of man (cf. 20). Through successive generations of interbreeding, it has been argued, mixed types have been produced which now outnumber the remaining specimens of pure types. It is well known that, through the mechanism of heredity, interbreeding will in the long run produce a larger number of mixed than pure individuals. The same could apply to interbreeding among the proposed human biotypes. This situation would then present a normal distribution of traits, with the largest number of individuals in the center of the distribution, corresponding to the numerically largest "mixed" group. Thus the form of the distribution curve cannot in itself indicate the composition of the group. The normal curve might be obtained with a single intermediate type and minor deviations from it, or it might result from the mixture of several pure biotypes.[9]

Constitutional Relationships. The only essential implication in the concept of "biotypes" seems to be a certain *organization* among the various characteristics of the individual. Thus a relationship would be expected between body build, emotional reactions, and intellectual traits. If there exist diverse biological types of man, each manifesting its own peculiarities in physique, personality, and intellect, we should find a certain degree of conformity among these characteristics of the individual. When so conceived, the problem of types is ultimately reducible to a consideration of the relationship between structural and

[9] For a more technical analysis of the logic of different typologies, cf. Winthrop (47, 48, 49).

behavioral qualities. It is not, however, concerned with isolated traits, but with the composite picture of the individual as a whole.

Methodology of Constitutional Studies. How can such theories of constitutional typology be tested? One method involves the classification of individuals into *extreme behavior groups* and the subsequent comparison of these groups in regard to physique. This technique has been employed largely with abnormal cases in the effort to check the assertions that a given physique predisposes the individual to a certain kind of mental disorder. Thus, for example, the relative number of pyknics and leptosomes among individuals manifesting different forms of insanity has been compared and evaluated in terms of the expected association.

A second method is based upon the *correlational analysis* of measurements collected on unselected normal groups. Various physical indices of body build have been worked out for this purpose. Such indices are then correlated with test scores or ratings on crucial personality traits. A high correlation would be evidence for the conformity implied by type theories.

In a few studies, which illustrate a third approach, efforts have been made to identify and select *"pure types"* on the basis of physical criteria; the psychological characteristics of the selected individuals are then thoroughly investigated. The subjects are originally chosen so as to represent "good specimens" of each type. These physically contrasted groups are then compared in emotional and intellectual reactions. This method is in a sense the opposite of the first method described above, which began with psychologically contrasted groups and proceeded to compare them in physique. The present method starts with groups clearly differentiated in physique and compares them in behavior. It should be noted, however, that the present method does not merely choose individuals who represent extremes in any one physical characteristic, such as height or weight. It is an essential feature of this method that individuals are chosen on the basis of a *composite* of physical specifications so as to fit the particular type pattern.

A fourth and more recently developed approach is to identify first the *basic components* of both physique and personality. These components constitute the categories in terms of which each individual is described or "typed" in both body build and behavior characteristics. The correlation between each individual's physical and psychological

status can then be found. This approach differs from the second one described above only in its emphasis upon *what* is to be correlated. The argument is that the correlations—or lack of correlations—heretofore found between physique and personality may be misleading because inadequate, superficial, or unessential aspects of both physique and personality were measured. The principal efforts of this approach are thus concentrated on discovering the basic categories in terms of which both the domain of body build and the domain of behavior can be described.

It should be noted that the differences among these four approaches are differences in emphasis rather than in long-range objectives. All are fundamentally concerned with the relationship between structural and behavioral characteristics. In the sections which follow, the investigations have been grouped under these four approaches primarily for convenience of presentation. Moreover, the order in which these four methods have been treated in the present section, as well as in the remainder of the chapter, is a chronological rather than a logical one. In terms of similarity of procedure, the first and third approaches might have been considered together, and the second and fourth could have been similarly grouped.

EVIDENCE FROM ABNORMAL CASES

Kretschmer originally formulated his theory of constitutional types from observations on psychotic patients. In comparing the body build of schizophrenics and manic-depressives, he consistently found a greater proportion of leptosomes among the former and pyknics among the latter. In one survey, Kretschmer (26) compiled data from several investigators on over 4000 abnormal cases, with the results shown in Table 19. It is apparent that by far the largest percentage of schizophrenics fall into the leptosome and athletic categories, and an equally large percentage of manic-depressives fall into the pyknic and mixed pyknic classes.

Wertheimer and Hesketh (46) measured 65 male patients chosen at random from two American institutions for the insane. Of these, 11 had been clearly diagnosed as manic-depressive and 23 as schizophrenic. The major part of the investigation was therefore confined to these cases. Such subjects were first classified into Kretschmer's body types on the basis of general observation. A series of 53 anthropo-

TABLE 19 *Per Cent of Schizophrenics and Manic-Depressives Falling into Different Categories of Body Type*

(From Kretschmer, 26, p. 34)

Body Type	*Schizophrenics*	*Manic-Depressives*
Pyknic and mixed pyknic	12.8	66.7
Leptosome and athletic	66.0	23.6
Dysplastic	11.3	0.4
Unclassifiable	9.9	9.3

metric measurements were then taken and various bodily indices computed. One of these indices was ultimately selected as the most satisfactory [10] and adopted as the chief basis of classification. A close correspondence was found between the two procedures. Those individuals classified as pyknic by the experimenter's diagnosis invariably had indices under 255; those classified as leptosomes had indices over 270. There was no overlapping in the indices of these two groups. By either method of classification, however, the number of decided pyknics or leptosomes was small, most individuals falling into the intermediate athletic or mixed groups, as would be expected.

TABLE 20 *Further Data on Per Cent of Schizophrenics and Manic-Depressives Falling into Different Categories of Body Type*

(From Wertheimer and Hesketh, 46, pp. 1040-1041)

Body Type	*Schizophrenics* (N = 23)	*Manic-Depressives* (N = 11)
Pyknic	4.3	45.5
Pyknoid	13.0	36.4
Athletic	26.1	9.0
Leptosome-athletic-mixed	34.8	0.
Leptosome	17.4	0
Unclear	4.3	9.0

The percentages of persons of each body type found in the schizophrenic and manic-depressive groups are given in Table 20. These data again show a marked predominance of pyknic types among the manic-depressives. The schizophrenics scatter over a wide variety of

[10] Index = $\frac{100 \times \text{leg length} \times 10^3}{\text{transverse chest diameter} \times \text{sagittal chest diameter} \times \text{trunk height}}$ (cf. 46, p. 415).

body type, but the greatest number fall into the leptosome and athletic groups.

The chief difficulty in interpreting the results of these and similar investigations on psychotic cases arises from an inadequate control of the *age factor*. Schizophrenia is more common among younger subjects, whereas older people are more susceptible to manic-depressive psychoses. It is also a well-established fact, which Kretschmer himself recognized, that older subjects tend more toward the pyknic body build, younger subjects toward the leptosome. To be sure, pyknics may be found among young people, and leptosomes among older groups; and many individuals retain the same type of body build throughout life. But the general trend is sufficiently marked to produce an entirely spurious relationship between body build and psychotic tendencies. For this reason, it is essential that age differences be ruled out in any comparison of the body type of different psychotic groups.

In an investigation by Garvey (14), 130 manic-depressives and 130 schizophrenics were selected so that the two groups were closely matched in age. Only clear cases, classified with complete agreement by the hospital staff (not including the experimenter), were employed. When the patients were divided into heavy and slender types on the basis of general observation, some evidence was found for Kretschmer's claims. The association, however, is reported as too slight to permit body type to be regarded as diagnostic of psychosis. Extensive physical measurements were taken and several ratios between horizontal and vertical bodily dimensions were computed. All showed an *almost complete overlapping* of the two psychotic groups. Not only were the averages closely similar, but also the range and the general form of the distribution were practically identical in the two groups.

Naccarati (31), in an effort to check upon Viola's hypothesis, measured 100 male Italian psychoneurotics between the ages of 25 and 40. The number of normosplanchnics is reported as being smaller in this group than in normal groups. The neurasthenics had a larger proportion of microsplanchnics (long, slender type), while macrosplanchnics predominated among the "emotional psychoneurotics." Under the latter category Naccarati included cases of hysteria, anxiety neuroses, and traumatic neuroses. Averages of some of the most significant physical measurements as well as the average age of the two groups are given in Table 21. It will be noted that the neurasthenic group has a lower average age than the "emotional psychoneu-

rotics." This might account in part for the greater tendency to microsplanchny among the former. No description is given of the method of obtaining or diagnosing the subjects, a fact which makes interpretation of the findings difficult.

TABLE 21 *Bodily Characteristics of Individuals with Different Forms of Psychoneuroses*

(From Naccarati, 31, p. 543)

Group	*Morphologic Index*	*Total Volume of Trunk*	*Length of Extremities*	*Age*
50 Neurasthenics	456.64	30.43	133.35	32.16
50 "Emotional psychoneurotics"	362.06	37.36	128.80	33.94

An extensive investigation on the relationship between body type and psychosis was conducted by Burchard (2). A total of 407 white male patients from several institutions for the insane were selected for the survey. Of these, 125 were clearly diagnosed as schizophrenes by the hospital staff, and 125 as manic-depressives. The remaining 157 patients manifested a variety of psychotic and neurotic conditions, and were employed as a control group. The subjects in all three groups were classified into pyknics, athletics, and leptosomes by "general impression." Comparisons were also subsequently made in respect to several anthropometric measures and indices. Only seven dysplastics were found in the entire sampling, and these were eliminated from further consideration. All other subjects were retained, any intermediate or mixed types being assigned to the morphological type which they resembled most closely. In Table 22 are given the percentages of pyknics, athletics, and leptosomes found in the manic-depressive, schizophrenic, and control groups, respectively, when the inspectional method of classification was employed.

The general trend of these figures seems to be in agreement with Kretschmer's theory. Not only are the greatest percentage of manic-depressives pyknic, and the greatest percentage of schizophrenes leptosome, but the control group occupies a position intermediate between these two groups in all percentages. When the schizophrenes and manic-depressives are compared in terms of anthropometric measures, a certain amount of differentiation is also revealed. Reliable

TABLE 22 ***Per Cent of Schizophrenic, Manic-Depressive, and Control Subjects Showing Pyknic, Athletic, and Leptosome Body Types***

(From Burchard, 2, p. 31)

Morphological Type	*Manic-Depressives*	*Schizophrenics*	*Control*
Pyknic	63.2	36.3	55.6
Athletic	8.8	17.7	11.3
Leptosome	28.0	46.0	33.1

differences between the averages of the two groups were found in three out of nine physical measures and in two out of three bodily indices. Nevertheless, the overlapping of the groups in all these meas-

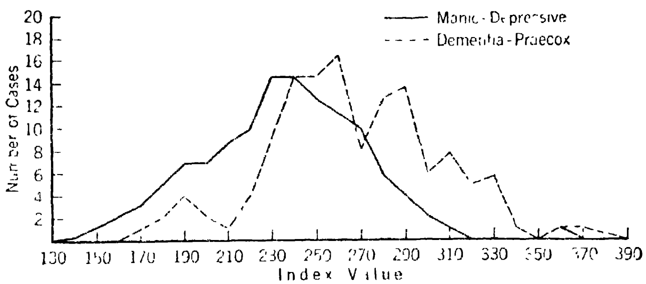


Fig. 71. Frequency Distribution of 125 Manic-Depressives and 125 Schizophrenes on the Wertheimer-Hesketh Index of Body Build. (From Burchard, 2, p. 47.)

ures was very large. This is illustrated in Figure 71, which shows the frequency distributions on the Wertheimer-Hesketh index of body build.[11] This index yielded the largest differences between the two groups. It is apparent that, despite the statistically significant differences in averages, schizophrenes can be found who are much more pyknic than certain manic-depressives, and vice versa.

Even the differences in averages between the two groups may be the result of other uncontrolled factor Burchard recognized this difficulty and undertook a detailed analysis of his manic-depressive and schizophrene groups. In regard to racial and national background, occupation, and educational status, no appreciable or consistent differences could be discovered. In age, however, the differences were

[11] Cf. footnote 10

very large, the average ages of schizophrenic, control, and manic-depressive groups being 30.97, 42.90, and 49.65 years, respectively. Further analysis revealed a definite relationship between age and body build. This factor seems to have accounted largely, although not entirely, for the group differences found.

Since the age factor plays such an important part in all studies on constitutional type, we may examine more closely Burchard's data on this problem. In Table 23 will be found the average Wertheimer-Hesketh indices of subjects falling in successive decades, within the entire sampling as well as within each psychotic group. These averages indicate a definite tendency toward a more pyknic body build with advancing age. This is manifested *within each psychotic group,* as well as in the entire group. Further corroboration of this finding is furnished by the correlation of −.256 obtained between age and index

TABLE 23 *Wertheimer-Hesketh Index in Relation to Age and Type of Psychosis*

(From Burchard, 2, p. 64)

	Mean Wertheimer-Hesketh Index			
Age	Entire Group	Schizophrenes	Manic-Depressives	Control
15–19	306.11	297.25	262.66	321.00
20–29	275.10	279.77	252.00	273.48
30–39	260.82	272.00	256.33	253.86
40–49	249.34	252.50	246.52	249.41
50–59	253.68	277.50	247.29	257.16
60–69	236.50	243.33	241.67	228.75

value in the entire sampling. Much of the difference observed between the two psychotic groups can therefore be attributed to age. It should be noted, however, that within each decade the schizophrenes have a higher average index than the manic-depressives. To be sure, the differences are considerably reduced by ruling out age, and the control group no longer retains its intermediate position, but a certain difference in the expected direction remains. This difference could possibly have resulted from other unsuspected factors in which the two psychotic groups may not have been equated. Or it may indicate an actual, although very slight, relationship between body build and type of psychosis.

CORRELATIONAL STUDIES WITH NORMAL GROUPS

It has frequently been objected that one cannot generalize from a slight correspondence between body build and certain forms of insanity to a relationship between structural and personality traits of normal individuals. The comparison of average values, furthermore, or of the percentage frequency of bodily types among different groups may exaggerate a very slight degree of association. Such comparisons tell us little about individual cases. For these reasons, a number of investigators have resorted to the correlation coefficient to obtain an exact quantitative measure of the amount of relationship within a group.

The correlation coefficient is affected not only by the *presence or absence* of clear-cut types within a group, but also by the *degree* to which a given typal characteristic is exhibited. This method seems to rest upon a slightly different principle than that underlying group comparisons. Thus if morphological index were found to correlate highly with intelligence, it would mean not only that the clearly microsplanchnic are more intelligent than the clearly macrosplanchnic, but also that, within the intervening range, the more microsplanchnic the individual, the more intelligent he will be. A lack of relationship between intelligence and body build within the intermediate mixed groups will considerably lower the correlation which would be obtained if only "pure types" were included.

Let us examine some facts and figures. Naccarati (30) found a correlation of .356 between morphologic index and intelligence test scores within a group of 75 college men. In the same study, the height-weight ratios [12] of 221 college men ranging in age from 17 to 22 correlated .230 with intelligence. At first glance, these slight positive correlations between height-weight ratio (or morphologic index) and intelligence would seem to support the claim that the tall, slender individual tends to be more intelligent. The age factor, however, must again be considered. Upon further statistical analysis of the data, it was discovered that the correlation of .230 resulted largely from a negative correlation between *weight* and *intelligence test score* within this group.[13] The more heavily built, stocky individuals at the age

[12] The height-weight ratio has frequently been substituted for the more elaborate morphologic index, for the sake of expediency, since the two indices are closely related. Naccarati (30), for example, found a correlation of .70 between the two in a group of 75 students, and a correlation of .75 in another group of 50.

[13] Subsequently computed by Hull (17, pp. 142-143), from Naccarati's published data.

levels under consideration tend to be the *older* members of the group. Similarly, the older individuals *within any one academic level* are usually the duller ones. It therefore seems very likely that even the low degree of correspondence found between height-weight ratio and intelligence is attributable to an uncontrolled age factor and cannot be accepted as proof of a relationship between body type and mentality.

In a subsequent investigation by Heidbreder (16), Naccarati's hypothesis was checked on a group of 1000 white, native-born college freshmen, including 500 men and 500 women. The correlations between intelligence test scores and height-weight ratios proved to be only .03 for the men and .04 for the women. Similarly, the correlations between height-weight ratios and scores on each of the five subtests of the intelligence examination closely approximated zero, ranging from −.07 to +.10.

In an effort to discover whether the use of the more elaborate morphologic index in place of the height-weight ratio might yield more positive evidence for Naccarati's view, Sheldon (36) conducted an intensive investigation on 434 freshman men, between the ages of 17 and 22. Twelve measurements were carefully made on each individual and from them was computed the morphologic index, in the manner described by Naccarati. The correlation between these indices and scores on a common group intelligence test for college freshmen was .14. Correlations of the morphologic index with each of the nine sub-tests in the examination ranged from −.02 to +.12. These findings corroborate closely those obtained by Heidbreder with the height-weight ratio.

In a further investigation of morphologic types, Sheldon (37) correlated morphologic index and ratings on five personality traits within a group of 155 freshman men. Each student was rated by five upperclassmen who belonged to the same fraternity as the subject. The judges had thus had considerable opportunity to observe the student's everyday behavior in many situations and were fairly well qualified to rate him. The consensus of all five judges was taken as the final rating for each individual. Below will be found the correlations between morphologic index and ratings on each trait:

Emotional excitability	.00
Aggressiveness	−.08
Leadership	−.14
Sociability	−.22
Perseverance	.01

On the whole, these correlations are too low to indicate an appreciable degree of relationship between bodily type and personality traits. The correlations of morphologic index with leadership and sociability are, however, suggestive. These two correlations indicate a tendency for the more heavily built individual to be more sociable and more of a leader. This could again result in part from an uncontrolled age factor, inasmuch as the older individuals within such a group might well be expected to manifest these characteristics.

A comprehensive investigation including both intellectual and personality traits was conducted by Garrett and Kellogg (13). The subjects were again male college freshmen. Morphologic indices were computed with measurements taken from three standard photographs of each subject. These photographs, taken in connection with gymnasium routine, showed three different views of the individual in the nude. The morphologic indices computed from the photographs correlated .81 with height-weight ratios obtained from direct measurements on 219 students. On this basis, the authors felt justified in their use of the photographs for the sake of expediency. The "photographic" morphologic indices, as well as the height-weight ratios from direct

TABLE 24 *Correlations of Height-Weight Ratio and Morphologic Index with Test Scores*

(From Garrett and Kellogg, 13, p. [illegible])

Test	*Morphologic Index (from Photographs)*		*Height-Weight Ratio (from direct measurements)*	
	Number of Cases	Correlation	Number of Cases	Correlation
Thorndike Intelligence Test	206	.07	204	.10
Woodworth P.D. Sheet	151	.05	150	.09
Social Intelligence Test	123	−.06	122	.05

measurements, were correlated with tests of intelligence, emotional stability, and social aptitude, with the results shown in Table 24. None of these correlations is sufficiently large to indicate a significant degree of relationship. Thus we must conclude that none of these correlational studies on unselected normal samplings has provided any support for constitutional typology.

THE STUDY OF "PURE TYPES"

Exponents of typology have been quick to object that the low, negligible correlations in unselected samplings could result from the presence of a large group of individuals of mixed types, in whom no consistent relationship between physique and personality may exist. These individuals, who are probably in the majority, would serve to "dilute" any clear-cut relationships among the "pure types." It has also been argued that even when indices are employed in lieu of isolated dimensions, the investigator is not getting a picture of the individual's physique in its totality. And the latter is essential in any concept of constitutional types.

Most of the numerous German investigations on types have proceeded by *selecting* good specimens of each type on the basis of physical measurements or observations and then administering a variety of psychological tests to the groups so obtained. By this method, for example, the conclusions were reached that pyknics are more distractable than leptosomes, that they have a greater perception span, show a better incidental memory, respond "synthetically" rather than "analytically" to a difficult perception, are more sensitive to colors than to forms, are superior in motor tasks except when these require delicacy of movement, and give more extroverted responses. These are among the major differences reported by German investigators.[14] These writers place relatively little stress upon differences in general intelligence between the types.

Many of these studies are open to serious criticism and it is therefore difficult to evaluate their findings. The groups employed were usually small. Averages were reported with no indication of variability within each group or of amount of overlapping between groups. Quantitative data were frequently lacking and only descriptive observations reported. The tests were often inadequate or poorly standardized. The groups themselves, selected chiefly on the basis of physical type, frequently differed in other essential respects. Thus the relative proportion of men and women may not have been constant in all the groups. Or the pyknics may have been older than the leptosomes, in which case this age difference could account for the observed psychological differences. Little or no attempt was made to control this age factor, in some studies the subjects ranging from adolescents to sexage-

[14] For a survey of many of these investigations, see Klineberg *et al* (20).

narians. Social and cultural background may also have affected the results. There is some evidence, for example, that leptosomes are found more commonly among the higher social and educational levels. Since there are also intellectual and possibly emotional differences from one socio-economic or educational level to another, such factors should be held constant.

In America, Mohr and Gundlach (29) conducted an intensive quantitative investigation on a group of male convicts in a state prison. A total of 600 men were measured, out of which 89 were selected as good representatives of leptosome, athletic, and pyknic types. In arriving at this classification, the investigators employed all the anthropometric measures suggested by Kretschmer, as well as a general observational diagnosis of body type. Each subject was then given the Army Alpha and about a dozen simple psychological tests suggested by the German workers as diagnostic of constitutional type. Included were such tests as speed of tapping and of writing, visual reaction time, cancellation, substitution, color fusion, and Rorschach inkblots.

A striking difference in average Alpha score was found among the three groups. This is shown in Table 25, together with the number of cases in each group and the average ages. The correlation between

TABLE 25 *Mean Intelligence Test Scores of Adult Leptosomes, Athletics, and Pyknics*

(Adapted from Mohr and Gundlach, 29, pp. 133, 134)

Body Type	*Number of Cases*	*Mean Age*	*Mean Alpha Score*
Leptosome	19	28.55	96.5
Athletic	26	28.65	79.2
Pyknic	44	34.75	57.9

Alpha score and an index of body build was found to be −.34, which further corroborates the above results. Although not very high, this correlation indicates a significant tendency for the tall, slender individuals to obtain higher scores. Similarly, in many of the other tests the differences among the groups were large enough to be statistically significant. It will be noted, however, that there is a marked difference in age among the three groups, the pyknics being on the average a little over *six years* older than the leptosomes or athletics. In view of

the tendency for Alpha scores to decrease with age (cf. Ch. 9), therefore, the pyknic group would be expected to obtain lower scores. The cultural and racial composition of the three groups is not stated, and this factor may also account for some of the observed differences in test performance.

In a later study, Klineberg, Asch, and Block (20) undertook to compare Kretschmer's body types under more rigidly controlled con-

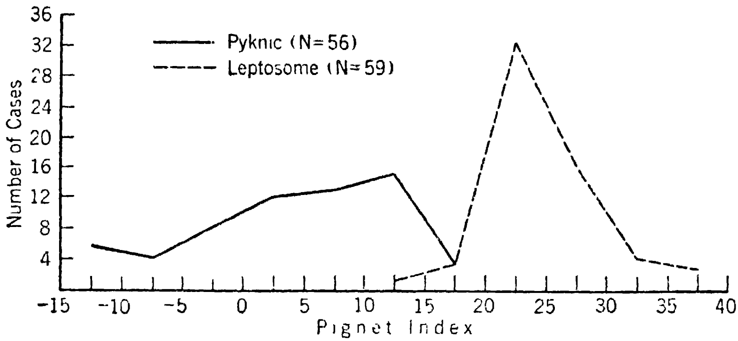


Fig. 72. Distribution of Scores of Leptosomes and Pyknics on the Pignet Index. (From Klineberg, Asch, and Block, 20, p. 180.)

ditions. The study was limited exclusively to college students, so that variations in age and in social and educational level were markedly reduced. The subjects were also very homogeneous in racial and cultural background. From a group of 153 men in a single college, averaging 19 years-9 months in age, it was possible to select 56 "pure pyknics" and 59 "pure leptosomes." The classification of body type was based upon five indices computed from physical measurements, together with the experimenter's observational diagnosis. That the two chosen groups were clearly differentiated in physique is illustrated by Figure 72. This shows the distributions of the pyknic and leptosome groups in Pignet Index,[15] one of the five criteria of selection employed. It will be noted that overlapping is virtually absent.

In sharp contrast to this distribution is that reproduced in Figure 73, showing the scores of leptosomes and pyknics on one of the

[15] Pignet Index = Height − (weight + chest circumference).

NOTE: This formula is printed incorrectly in the study under consideration (20, p. 164). We assume this was a misprint, and that the correct formula was employed in the computations.

psychological tests, viz., cancellation of letters. In this case, the two groups overlap almost completely. Similar results were obtained with all the other tests, which included tests of intelligence and of emotional adjustment, as well as six tests specifically designed to measure alleged characteristics of the two opposed constitutional types. In no case were the differences between the two groups statistically significant. Correlation of measures on 110 cases confirmed these findings. The

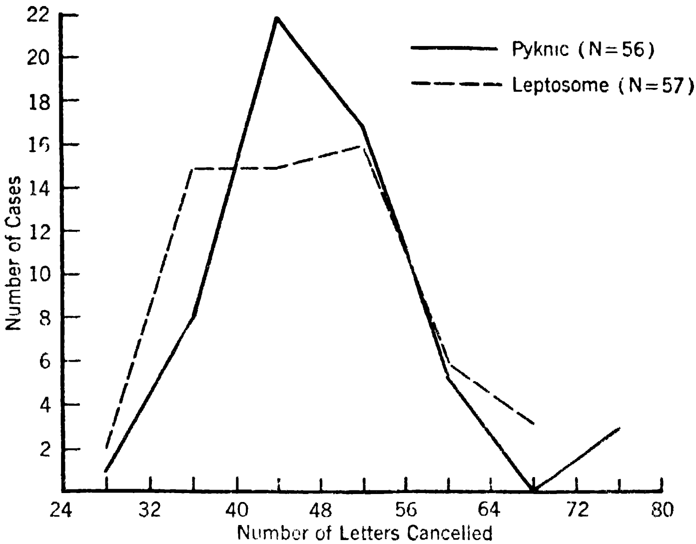


Fig. 73. Distribution of Scores of Leptosomes and Pyknics in Letter Cancellation. (From Klineberg, Asch, and Block, 20, p. 180.)

correlations between physical indices and test scores were all close to zero. Intercorrelation of the various psychological tests were also negligible. If the underlying conformity implied by type theories were present, a fairly close correspondence should have been found among the various diagnostic tests. Viewed from any angle, the results are completely negative.[16]

An intensive investigation of personality traits in relation to physical type was conducted by Klineberg, Fjeld, and Foley (21). The

[16] A parallel study was conducted on 79 women students, but the results were less conclusive since all the women fell within the leptosome range and no genuine pyknics could be found. Comparisons within the leptosome group, both by correlational and by contrasted group methods, corroborated the findings on the male students.

subjects were again students, selected from several colleges in New York City and its environs. A total of 200 men and 229 women were examined. Within each of these groups, the subjects who fell in the upper and those who fell in the lower 25% of the distribution of Pignet Index were selected as leptosomes and pyknics, respectively. This gave 50 leptosomes and 50 pyknics among the men, and 57 leptosomes and 57 pyknics among the women. These contrasted physical types also showed significant differences in nearly all other physical measures and indices obtained in the study, and can safely be regarded as distinctly different in physique. Little or no overlapping was found in any of these measures.

As for age, the male pyknics proved to be slightly older than the leptosomes, their average ages being 21.08 and 20.17, respectively. Besides being very slight, this age difference is such as to *exaggerate* any of the alleged psychological differences between leptosomes and pyknics. Hence such an age discrepancy loads the dice slightly in favor of Kretschmer's hypothesis and would make *negative* findings all the more conclusive. Among the women, the age difference was negligible, the leptosomes averaging 19.73 and the pyknics 19.23 years.

All subjects were given the Bernreuter Personality Inventory, the Allport-Vernon Study of Values, and a specially designed test of suggestibility. A large portion of the group also took two other tests. One of these was an honesty test, showing the number of times the subject cheated on what seemed to be an information test (Maller Test of Sports and Hobbies). The other was a specially constructed persistence test which measured the length of time the individual worked on an insoluble finger maze before giving up. In Table 26 will be found the average scores of both male and female leptosomes and pyknics on each test, together with data on the significance of the differences between the averages.

The results of this carefully controlled study are clearly negative as regards type theories. None of the leptosome-pyknic differences, in either male or female group, is significant. In other words, all the obtained differences could have resulted from chance errors of sampling. It should also be observed that in several comparisons the differences between leptosomes and pyknics were contrary to expectation. For example, the male leptosome group appears more "sociable" on the Bernreuter and seems to have a higher sense of "social value" accord-

TABLE 26 *Average Scores of Leptosome and Pyknic Groups on Personality Tests*

(From Klineberg, Asch, and Block, 22)

Test	*Male*			*Female*		
	Leptosome (N = 50)	Pyknic (N = 50)	diff./σ_{diff}	Leptosome (N = 57)	Pyknic (N = 57)	diff./σ_{diff}
Bernreuter						
1. B_1N: Neuroticism	−37.50	−39.76	0.14	−40.51	−45.60	0.34
2. B_2S: Self-sufficiency	+34.30	−29.22	0.49	− 0.26	−18.02	1.88
3. B_3I: Introversion	−14.92	−17.88	0.30	−18.47	−27.00	0.95
4. B_4D: Dominance	+39.20	+42.18	0.26	−30.32	+38.54	0.72
5. F_1C: Self-confidence	− 8.51	− 8.46	0.005	+ 4.67	−13.67	1.19
6. F_2S: Sociability	+ 1.22	− 6.66	0.62	−29.79	−18.39	0.97
Allport-Vernon Study of Values						
1. Theoretical	31.82	31.46	0.23	28.57	29.32	0.54
2. Economic	28.67	29.98	0.96	27.42	26.71	0.61
3. Aesthetic	27.27	28.40	0.61	34.77	32.91	1.15
4. Social	32.29	31.82	0.39	30.25	32.63	2.22
5. Political	30.30	31.72	1.08	29.65	29.54	0.09
6. Religious	29.65	26.62	1.47	29.34	28.89	0.28
Suggestibility	11.02	10.82	0.23	11.89	11.43	0.56
Persistence *	6.09	10.94	2.22	6.94	8.71	1.32
Honesty *	97.58	94.77	1.61	99.00	99.14	0.25

* Not all subjects were given these tests.

ing to the Allport-Vernon scale than does the male pyknic group. The average scores of all the groups, furthermore, fell very close to the norms for college men and women in general. A final point to note is that the ranges of the leptosome and pyknic groups were nearly identical, showing an almost complete overlapping of distributions on all personality tests.

THE SEARCH FOR COMPONENTS OF PHYSIQUE AND TEMPERAMENT

The approach illustrated by the recent contributions of Sheldon and his co-workers (39, 40) emphasizes the discovery of *basic components* of "physique" and "temperament." This is regarded as a necessary first step in any investigation of the relationship between structural and behavioral characteristics. In Sheldon's schema for classifying physiques, the individual is rated on a 7-point scale in each of the three following components:

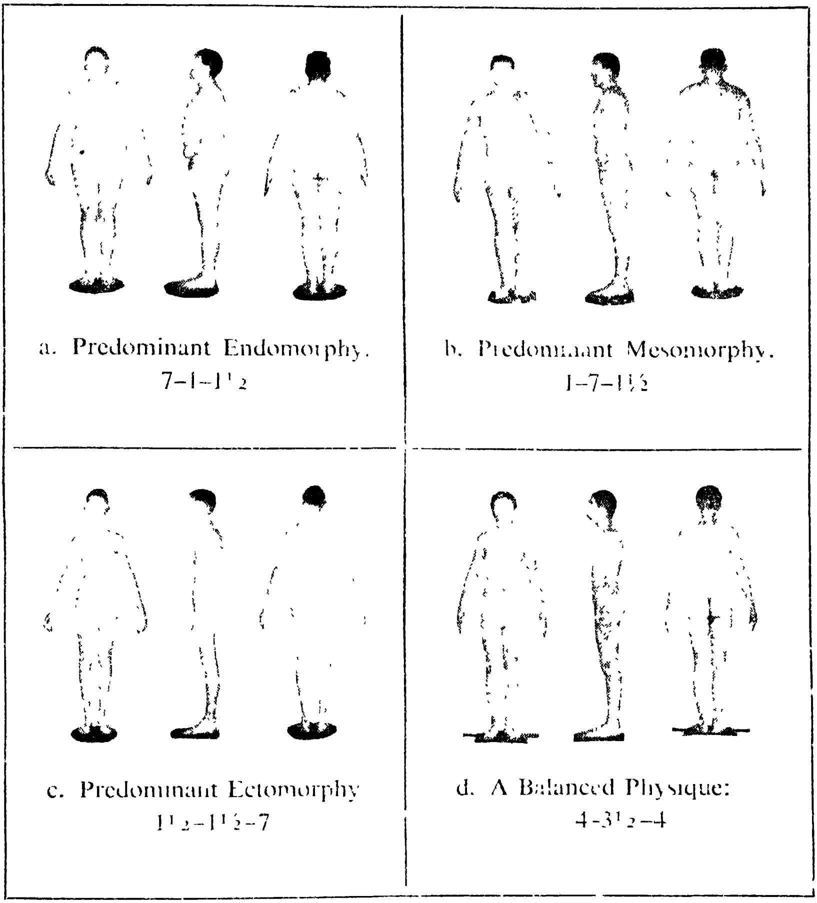
(1) *Endomorphy*—the degree to which "soft roundness" predominates. A person rated high (at or near 7) in endomorphy is flabby, soft, and rolypoly. In such a physique, the digestive viscera are overdeveloped in relation to other body systems, and the individual has a relatively large abdominal and thoracic cavity.

(2) *Mesomorphy*—the degree to which bone and muscle predominate. The professional "strong man" of the circus would usually rate 7 in this component. The distinguishing mark of mesomorphy is uprightness and sturdiness of structure.

(3) *Ectomorphy*—the degree to which linearity and fragility predominate. The extreme ectomorph is "skinny," with long, delicate bones and underdeveloped muscles.

Each individual's "somatotype" consists of three numbers, representing his rating in endomorphy, mesomorphy, and ectomorphy, respectively. Thus a 7–1–1 represents extreme endomorphy. A 2–6–2 and a 3–6–2 are both highly mesomorphic, but the latter shows more endomorphy than the former. Theoretically, there are 210 somatotype combinations which could be obtained with three components rated on a 7-point scale. But some of these combinations are obviously impossible, such as the hypothetical 7–7–7 or 1–1–1. Sheldon (39) describes 76 somatotypes which have been actually observed. In Fig-

ure 74 will be found photographs of four somatotypes, illustrating the extremes of endomorphy (7–1–1½), mesomorphy (1–7–1½), and ectomorphy (1½–1½–7), and one physique in which the three com-



a. Predominant Endomorphy. 7–1–1½

b. Predominant Mesomorphy. 1–7–1½

c. Predominant Ectomorphy 1½–1½–7

d. A Balanced Physique: 4–3½–4

Fig. 74. Human Somatotypes. (From Sheldon and Stevens, 39, pp. 8-9.)

ponents are nearly in balance (4–3½–4). The most common physique among the college men who were Sheldon's principal subjects was 3-4-4.

One of the principal advantages of this system of classification over earlier constitutional typologies is that it clearly begins with the as-

sumption of a *continuous distribution,* and merely describes the individual in terms of three components or variables rather than one. The use of a 7-point scale is of course arbitrary and only a matter of convenience. Five, ten, or any other number of steps could be substituted, in which case the total number of somatotypes would be decreased or increased. A further advantage which the authors claim for this system is that nutritional and age differences do not alter the individual's somatotype. For example, they maintain that although most endomorphs are "fat," loss of weight will not change endomorphs into mesomorphs or ectomorphs—"they become simply emaciated endomorphs" (39, p. 8). The reason for this consistency of somatotype is that several measurements are taken in parts of the body which change little when flesh is added or lost. In somatotyping an individual, the three components are rated in at least five different bodily regions and then averaged. For example, the separate ratings of the 7–1–1½ pictured in Figure 74a were 7–1–2, 7–1–1, 7–1–2, 7–1–2, 7–1–1.

The three components of physique were chosen and defined through detailed observations of the photographs of 4000 college men. As shown in Figure 74, each person was photographed in the nude and in a standardized posture, from the frontal, lateral, and dorsal positions. Once the three components had been suggested by this inspectional analysis, suitable anthropometric measures were selected by trial and error. The measurements were made directly on the photographs with needle-point dividers.

A parallel procedure was followed in arriving at the basic components of temperament. First, the authors assembled a list of 650 alleged temperamental "traits" described in the literature, most of them being related to introversion-extroversion. After adding a few from their own observations, arranging, and condensing, they were able to reduce the list to 50 terms which seemed to embody all the essential characteristics. A group of 33 young men, mostly graduate students and instructors, were then rated on a 7-point scale for each of these 50 traits. The ratings were based upon a series of 20 intensive interviews by the experimenter, extending over a period of one year and supplemented by everyday observations. All the 1225 intercorrelations among the ratings of the 50 traits were computed. An inspection of this correlation table suggested to the authors that the traits fell into three principal "clusters," such that the tests within each cluster were positively correlated with each other and negatively cor-

related with the tests in the other clusters. At this point it was decided to keep only those traits, or items, which had a positive correlation of .60 or more with other items within their cluster and a negative correlation of .30 or more with items outside the cluster. On this basis 22 of the original 50 traits were retained. In the course of subsequent studies on more subjects, the investigators undertook to sharpen and redefine the initial 22 traits and to add others which also satisfied the above correlational criterion. The final scale developed by this technique consisted of 60 traits, 20 in each cluster.

The temperamental components represented by each of these three clusters were described as follows:

(1) *Viscerotonia*—tendency toward relaxation, love of comfort, sociability, conviviality, pleasure in eating and in digesting, indiscriminate amiability, complacent tolerance, easy emotional expression, and need of people when troubled.
(2) *Somatotonia*—tendency toward assertiveness in posture and movement, energetic activity, love of power and risk, physical courage, directness of manner, psychological callousness, general noisiness, and need of action when troubled. Individuals high in this trait are characterized by "vigor and push."
(3) *Cerebrotonia*—tendency toward restraint and tightness in posture and movement, love of privacy, fear of people, emotional restraint, poor sleep habits, and need of solitude when troubled.

How are the components of physique related to the components of temperament in this system? Sheldon maintains that endomorphy tends to be associated with viscerotonia, mesomorphy with somatotonia, and ectomorphy with cerebrotonia. In a group of 200 university men between the ages of 17 and 31, observed by Sheldon over a 5-year period, the following correlations were found between ratings for the individual components of physique and temperament:

Endomorphy and viscerotonia	.79
Mesomorphy and somatotonia	.82
Ectomorphy and cerebrotonia	.83

From a further analysis of the same subjects, the authors suggest the hypothesis that certain *discrepancies* between somatotype and temperamental index may predispose the individual to maladjustment and interfere with his achievement (39, Ch. VII).

The correlations between structural and temperamental components reported by Sheldon are certainly much higher than those found

heretofore. Sheldon and his co-workers attribute this difference to their own reliance upon "essential underlying components" of both physique and temperament, in place of what they regard as the relatively superficial or fragmentary measures of earlier investigators. Sheldon argues, for example, that aptitude or personality tests may not reveal the "deeper and more enduring aspects" of temperament which he claims to have reached through his series of interviews (7, p. 33). For this reason, test scores might not yield such high correlations with somatotype as were found by Sheldon through the use of ratings. A counter-argument is that the well-known "halo effect" may have operated in the ratings, producing artificially high correspondences between physique and temperamental index. As the strongest defense against the halo effect, Sheldon offers the fact that the experimenter was aware of its nature and was therefore on guard against it. The effectiveness of such a safeguard is of course debatable.

Subsequent checks on Sheldon's theory have shown that when objective test scores were substituted for personality ratings, the correlations between somatotype and behavior characteristics dropped to the familiar low values. In a study by Child and Sheldon (7) on groups of Harvard men,[17] somatotype was correlated with tests of verbal and numerical aptitude, ascendance-submission, and masculinity-femininity. These correlations were uniformly low, the highest being .21. Only a few were of marginal significance, indicating a fair probability of a true but very slight relationship. Fiske (11) found no significant association between Sheldon's somatotype classification and a series of intelligence and personality tests among 133 13- to 17-year-old boys. In this study, ratings of personality characteristics also failed to show a significant relationship to body build.

Another serious criticism against the system proposed by Sheldon concerns the original identification of the three temperamental components (cf. 1). In the last analysis, the entire structure of evidence for the presence of these particular components stands or falls with the adequacy of the initial experiment on 33 Harvard men. To be sure, subsequent studies were conducted on larger groups. But these studies were designed simply to redefine, sharpen, and expand the originally chosen list of 22 "traits" for measuring the three temperamental components, rather than to check the adequacy of the compo-

[17] The number of subjects used for obtaining the different correlations varied from 90 to 518.

nents themselves. This is clearly indicated by the authors' procedure. The criterion for adding a new trait to the list was that the trait must correlate highly and positively with the traits in one of the original clusters, and negatively with the traits in the other two clusters. The subsequent modification or addition of traits thus depended in a very intimate way upon the results of the initial experiment. The small number and highly unrepresentative nature of the subjects employed in this initial experiment make it ill-suited to play such a fundamental part in the development of the entire schema of temperament classification.

Finally, the technique of identifying components by *inspection* of a correlation table leaves too much to subjective judgment. In so far as the major contribution of Sheldon's approach is its emphasis upon components rather than types, the best available objective techniques for identifying such components ought to be applied. These techniques, known collectively as "factor analysis," are based upon further statistical analysis of the table of intercorrelations, and will be discussed more fully in Chapter 15. In the present connection it will suffice to note that other investigators have begun to employ these techniques in the analysis of body build as well as personality characteristics, with results which offer little support to Sheldon's tripartite classification.

Cyril Burt (3, 4, 5), who first applied factor analysis to body build, has studied several groups differing in age, sex, and national background. The largest of these samplings consisted of 2400 British airmen in the RAF. Factor analyses were conducted on as many as 17 bodily dimensions, although some of the studies employed fewer measures. Such investigations by Burt and his students have indicated the presence of a general "size" factor, present to some degree in all bodily measurements and varying independently of type of physique. A second, bipolar or "type" factor was also found, which appeared to be related to the tendency toward a linear or a broad body build. This factor was positively related to measures of length, and negatively related to breadth, depth, and circumferential measures.[18] An index of body build based upon this second, bipolar, factor was found

[18] In a more detailed and refined factor analysis of some of these data, Thurstone (44) identified seven different factors, but these offer no more support to the Sheldon classification than do the original factors identified by Burt and his students. For other applications of factor analysis to body build, cf. 3, 4, 5, 15, 33, 34, 43.

to be normally distributed and to exhibit no tendency toward distinct types. Preliminary exploration of the relationship between this index and personality characteristics in both normal and abnormal groups showed a very slight but consistent trend along the lines suggested by Kretschmer and others (3, 34). Thus the "long-lean" person tended to be more inhibited and repressed, and more prone to schizophrenia and allied disorders; the more "rotund" individual inclined more toward cheerful, sociable reactions and toward cyclical forms of insanity. But again it should be repeated that these relationships were so slight as to be of theoretical interest only.

CONSTITUTIONAL TYPE OR SOCIAL STEREOTYPE?

Although too low to be of practical diagnostic value, the correlations between body build and certain personality characteristics seem to be sufficiently consistent to merit some consideration. Do these correlations mean that common, underlying "constitutional factors" exert at least a modicum of influence upon the development of both behavior and physique? Perhaps, but there is also another explanation. The differences in personality may have developed as a response to the differential treatment accorded to individuals of different physique by their social environment (cf. 6, 7). The existence of social stereotypes creates a vicious circle which tends to perpetuate whatever beliefs may be current regarding the association of physical and psychological traits.

Some of the effects of the individual's physique upon his social environment may be more direct and need not imply social stereotypes. This is especially true in childhood, when physical size, muscular prowess and strength, agility, and other characteristics associated with physique may influence social interactions and status. It has been repeatedly shown, for example, that adolescent leaders tend to be taller and heavier than their associates (cf. 8, pp. 248–249). Moreover, high school leaders are more likely to become leaders in college than those who were not leaders in high school, and the same individuals are also more likely to be community leaders after leaving school (9). It will be recalled, also, that in a previously cited study on fraternity men (cf. p. 438), the highest correlations found with morphological index—and the only ones approximating statistical significance—were those of ratings for leadership and sociability. These

correlations showed a tendency for the more heavily built man to be regarded as more sociable and more of a leader by his fraternity brothers. In the more recent study by Child and Sheldon (7), discussed in the preceding section, one of the highest correlations was between dominance scores and mesomorphy. High ratings in mesomorphy indicate a tendency toward overdevelopment of muscle and bone, suggesting the sturdy, athletic physique.

In an investigation designed to test the Kretschmer theory, Cabot (6) obtained an extensive series of test scores, ratings, and interview data on 212 high school boys. Within this group, 9 boys were clearly and consistently classified as pyknics, 25 as athletics, and 28 as leptosomes. The principal comparisons were made among these three groups. An examination of the personality measures showed that the largest differences occurred between the athletics, on the one hand, and the pyknics and leptosomes, on the other. The athletics tended to be more dominant and extroverted, and more active as social leaders. They were also *rated* higher in creativeness, imagination, responsibility, and influence on their associates. On this basis, Cabot proposed a theory of "socio-biological advantage," according to which a "good" physique (e.g., athletic or mesomorphic) gives the individual certain advantages in his social environment. It should be noted, in further explanation of Cabot's results, that a fairly significant difference in socio-economic level also favored the athletic group in his sampling.

In summary, then, the slight correlations found between body build and psychological characteristics in various studies are of the order of all other structural-behavioral correlations reported in the preceding chapter. And they are subject to the same variety of explanations. We have seen that, beginning with the sharply differentiated classifications of traditional type theories, typologists progressed through the "constitutional" concepts of relationships between structural and behavioral patterns, to the underlying question of the organization of all traits within the individual. It is to this basic question that we shall direct ourselves in the two chapters which follow.

CHAPTER

14

Variability within the Individual

THE STUDY OF VARIATIONS from trait to trait within the individual is of both practical importance and theoretical significance. When a child is classified as intellectually inferior on the basis of, let us say, Stanford-Binet IQ, there is still much that remains to be known about his abilities. Is he equally inferior in all respects, or does he exhibit significant discrepancies in his mental development? Is he normal or even superior along some specific lines? Similarly, in the case of a child of very high IQ, we may inquire in what ways he is superior. How uniformly does he excel the average child in intellectual performance? The intelligence test, furnishing a single summary figure to characterize the child's general mental level, often obscures important facts. Two individuals obtaining the same total score may present very different "mental pictures" when their performance along specific lines is analyzed.

The experienced psychologist has always taken this into consideration in interpreting test scores. The child's performance on the different parts of an intelligence scale and even, when feasible, on several different kinds of tests, is carefully analyzed before a final judgment is offered. There is a growing realization, however, that the question of variation among the individual's abilities deserves serious and systematic consideration and should be investigated in its own right. This problem is gradually coming to be regarded as even more important than the establishment of the individual's general level of performance.

In planning an educational program for a given individual, in helping him to choose a vocation, or in evaluating his qualifications for a job in industry, it is of the greatest importance to know his strong and his weak points, his particular assets and liabilities. Total

scores on intelligence tests can be used only in a crude and general sort of educational and vocational guidance. In the comparative study of groups, such as the sexes or different racial or cultural groups, a consideration of the general level of ability may also prove misleading. Let us suppose, for example, that one such group excels markedly in ability A and the other in ability B. If both are examined with an intelligence test which samples abilities A and B to an equal extent, no difference in total score will appear between the two groups. Essential and large differences might thus be concealed by the practice of lumping a number of tasks indiscriminately in the effort to arrive at the general mental average called "intelligence."

Much confusion has likewise been introduced into the interpretation of test results by the common tendency to accept labels too literally. Thus, if two tests are labeled measures of "intelligence," it is often incorrectly assumed that they are measuring the same characteristic of the individual. It is therefore most disconcerting to discover that the same child may appear dull on one intelligence test and above average on another. Such cases are, however, found. Since intelligence scales consist of a more or less random sampling of different tasks, the specific abilities covered by the various tests may differ. Some tests, for example, may be more heavily "loaded" with verbal material, others with spatial material. Even successive levels of the same test occasionally involve different abilities. Thus the Stanford-Binet draws more heavily from the verbal field at the higher year levels than it does at the lower. The same child tested with the Stanford-Binet at different ages might be favored at one time and handicapped at another because of the particular abilities called into play.

If the individual's abilities were all more or less on a dead level, a single summary score would be quite informative. But if appreciable variation in the individual's standing in different traits is the rule, then such a score is crude at best and may upon occasion be definitely misleading. It is essential, therefore, to inquire into the extent of variation within the individual. The data on this question have been gathered from a variety of sources. Individuals who exhibit *marked asymmetry of development* along different lines have been studied. Such individuals can be found among the feebleminded and the intellectually superior, as well as among the normal. Quantitative measurements have also been made on the extent of *variability from trait to trait* in large random samplings. Relevant data have likewise been

provided by *correlational studies.* An examination of the intercorrelations of scores on various tests indicates the degree of correspondence between the individual's performance along different lines.

THE PROFILE ANALYSIS

In the effort to obtain a more objective and concrete picture of variations within the individual than is furnished by the general impression of the examiner, a *psychograph,* or *profile chart,* of the individual may be drawn up. The psychograph shows at a glance the relative standing of the subject on any number of tests or other measures. The individual's scores on all tests must first be transmuted into *comparable units.* This is the fundamental step in any attempt to study variations within the individual. The psychograph itself, in the sense of a pictorial representation, could easily be dispensed with. The same information, although in a less vivid form, could be got from an examination of a set of scores obtained by the individual, *provided that all scores are expressed in the same terms.* It is in this latter respect that the judgment of the examiner needs to be supplemented by quantitative techniques. Confronted with a set of scores, some of which are expressed in seconds, others as number of words recalled, and still others as number of problems correctly solved, the examiner can tell little or nothing about the individual's relative standing in different functions.

Scores in different tests can be made comparable by the use of any of the common types of norms described in Chapter 2. If all the tests have been standardized in terms of age, each test score can be expressed as a *mental age.* For purposes of profile analysis, it is not necessary to convert the MA's to IQ's, since the CA is constant for any one individual. The use of age norms, however, is not always feasible. Some tests, especially in the field of personality, fail to show large or systematic age changes. In such cases, the range of variation within a single age group may be greater than the largest difference between the averages of different age groups. The application of the mental age concept to adults is also a questionable practice.

A more generally applicable technique is to convert the individual's score on each test into a *percentile.* The percentile, it will be recalled, indicates what per cent of persons are surpassed by the given individual (cf. Ch. 2). Thus if the subject falls at the 79th percentile

in arithmetic reasoning, we know that in this respect he excels a little more than three-fourths of the standardization group. Were the same individual to receive a percentile score of 34 in a vocabulary test, we would conclude that his performance in this function is considerably inferior to that in arithmetic, since only 34% of the standardization group falls below his vocabulary score. The results of different tests can also be made comparable by the use of *standard scores* (cf. Ch. 2). In this case, the subject's score is expressed as a deviation above or below the average of the standardization group, the unit being the standard deviation (SD) of the same group. Thus if his raw score falls exactly at the average, he receives a standard score of zero. A standard score of +1.00 signifies that the subject excels the average by one SD, and a standard score of —.5, that he falls short of the average by one-half of the SD.

It should be borne in mind that none of these techniques for converting scores into comparable measures yields a scale of equal units. They simply express, in terms which are more or less intelligible, the *relative position* of the individual in different tests, but they do not furnish a precise statement of the actual *amount* of trait difference represented by the various scores. Thus it will be recalled that the mental age unit corresponds to the average change in score occurring during a one-year period. Successive mental ages will not, therefore, represent equal increments of ability. We know that an MA of 6 indicates a higher standing than an MA of 5, and that an MA of 10 indicates a higher standing than one of 9; but we cannot conclude that the amount of difference is the same in both instances. Moreover, a change from a mental age of 5 to a mental age of 6 does not represent the same amount of improvement on two different tests, unless scores on the two tests show identical progress with age.

Nor can percentile scores be interpreted as equal ability units. As was shown in Chapter 3, such an interpretation would imply that the distribution of the trait measured is rectangular. Since, however, the distributions actually obtained conform more nearly to the normal bell-shaped curve, individuals will cluster more closely at the center of the distribution and scatter as the extremes are approached. Consequently, a difference of one percentile point at the extremes corresponds to a much greater difference in amount of the trait than does a difference of one percentile point nearer the center. The difference between an individual who receives a percentile rating of 90 in

height and one who receives a percentile rating of 91 is much greater, *in actual inches,* than the difference between two individuals receiving percentile ratings of 50 and 51.

Similarly, standard scores do not represent equal units. By subtracting a constant (the average) and dividing by a constant (the SD), we do not alter the scores in any essential way. The set of meas-

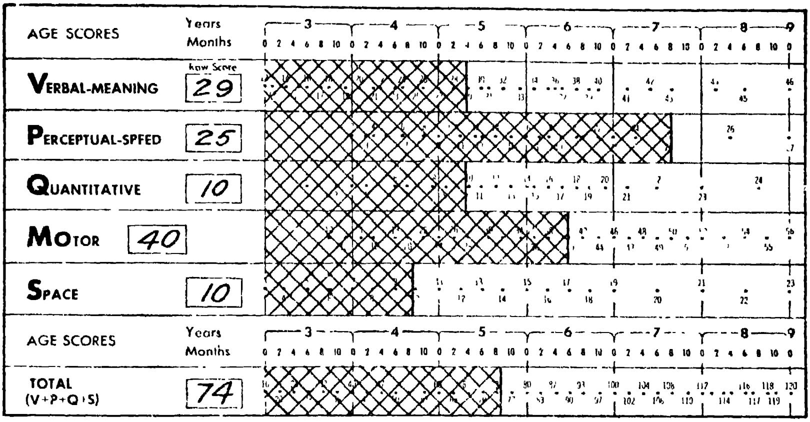


Fig. 75. Profile of a Five-Year-Old Child on the Thurstone Tests of Primary Mental Abilities. (Unpubl. data by courtesy of Miss Sonia Avakian.)

ures is simply transmuted into a different system of expression, as when pounds are changed to kilograms. The standard scores so obtained retain any inaccuracies or inequalities which were present in the original measures.[1]

Once the individual's scores on different tests have been expressed in the same terms, his profile can be plotted. A number of such profiles, obtained in a wide variety of testing situations, are illustrated in Figures 75 to 82. Figure 75 is a mental-age profile showing the relative standing of a 5-year-old kindergarten child in a number of functions commonly included in intelligence tests and measured by the Thurstone Tests of Primary Mental Abilities (47). This particular child is of about average intelligence, as indicated by her composite

[1] Only when scores on all the tests under consideration are normally distributed will standard scores represent equal amounts of ability in the different tests. Under these circumstances, the standard scores are identical to the scaled T-scores found by reference to normal curve frequencies.

IQ of 103. Her profile, however, shows a range of performance from 4 years-9 months in the space tests to 7 years-8 months in perceptual speed. In motor performance she is about one year accelerated over her age norm, while her mastery of verbal meaning and quantitative concepts is about normal for her age.

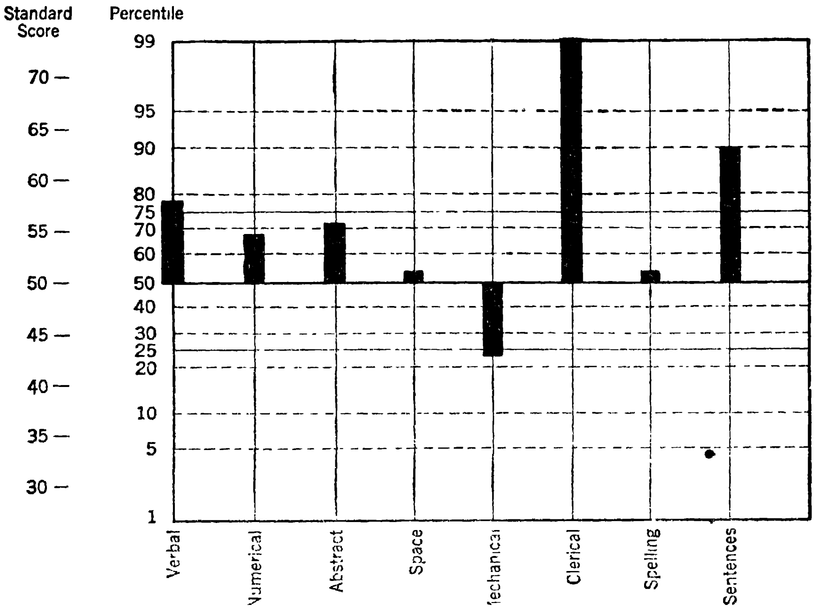


Fig. 76. Profile of a High School Boy on the Differential Aptitude Tests. (From Bennett, Seashore, and Wesman, 4, p. E8.)

In Figure 76 will be found a percentile profile of a high school boy on the Differential Aptitude Tests prepared by The Psychological Corporation (4). This student, whose IQ is 115, had planned to enter an engineering school. His mediocre performance on the test of space relations and his poor showing in mechanical comprehension raise doubts about his qualifications for engineering. On the basis of his other scores on the battery and his academic record, the student was advised to enroll in a general college course and to postpone decision regarding a field of specialization. An additional feature of interest in this type of profile chart is the spacing of percentile values in such a way that distances along the chart correspond to equal units of

ability in a normally distributed group. Thus, for example, the distance between the 80th and 90th percentiles is much greater than that between the 50th and 60th, as would be expected in a normal distri-

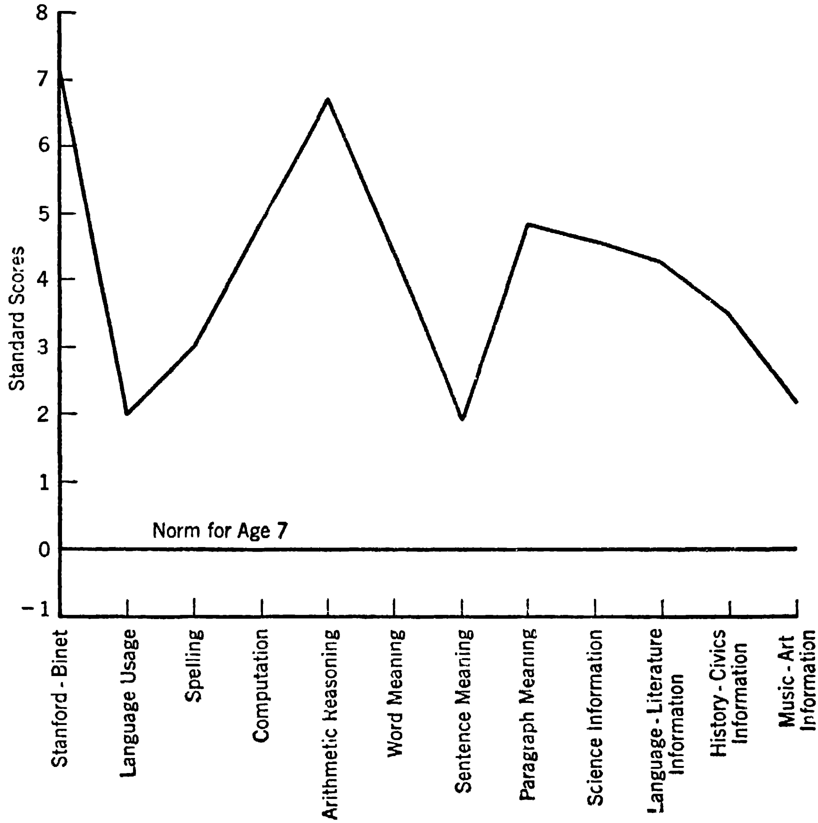


Fig. 77. Profile of an Intellectually Superior Seven-Year-Old School Boy. (From DeVoss, 10, p. 351.)

bution. Converted standard score equivalents are also shown to the left of the percentile scale.

Figures 77 and 78 are standard-score profiles of two school children selected for their outstanding intellectual level.[2] The psychograph in Figure 77 is that of a boy with a Stanford-Binet IQ of 173. This child,

[2] Part of the group of gifted children studied by Terman and his associates at Stanford University (cf. Ch. 17).

although above his age norm in all the tests, exhibits marked discrepancies among his scores. He is highest in arithmetic reasoning and computation. In the first of these he stands about 6¾ SD's above

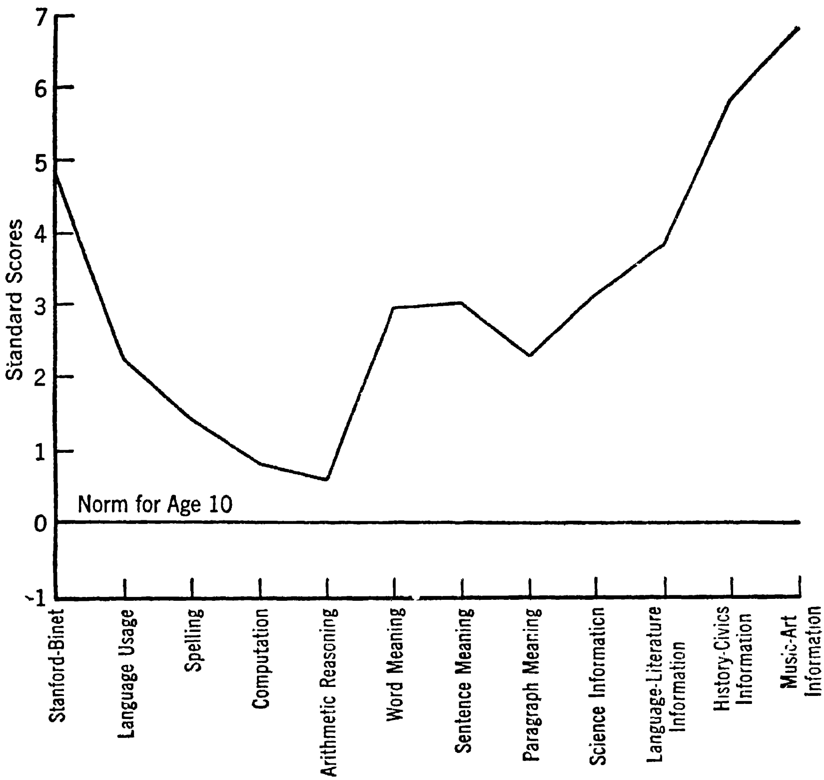


Fig. 78. Profile of an Intellectually Superior Ten-Year-Old School Boy. (From DeVoss, 10, p. 360.)

his age norm, and in the second, nearly 5 SD's. His performance is poorest on two of the verbal tests, language usage and sentence meaning. In Figure 78 is the psychograph of a school boy with an IQ of 155, who presents a very different picture. He is best in music-art information, second best in history-civics information, and poorest in arithmetic reasoning and computation. These examples illustrate the fact that intellectually superior children, although above their age

norms in most mental tests, may be much farther above average in some traits than in others.

The use of the profile technique in the investigation of occupa-

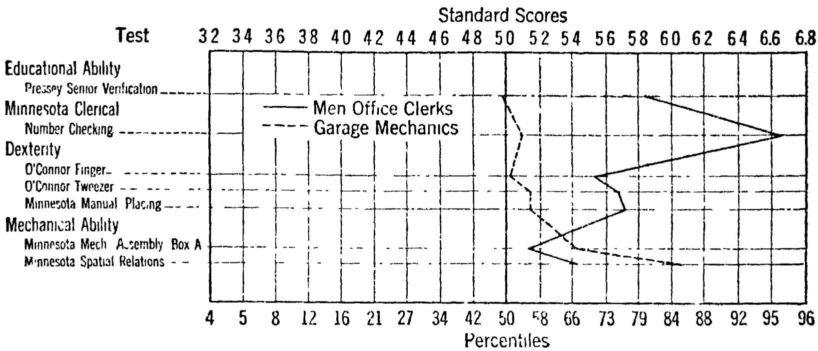


Fig. 79. Occupational Profiles of Men Office Clerks and Garage Mechanics. (From Dvorak, 11, p. 12.)

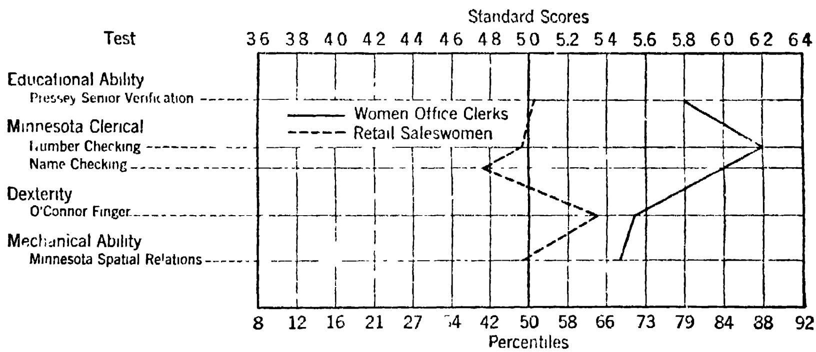


Fig. 80. Occupational Profiles of Women Office Clerks and Retail Saleswomen. (From Dvorak, 11, p. 16.)

tional ability patterns is illustrated in Figures 79 and 80, based upon data collected by the University of Minnesota Employment Stabilization Research Institute (11). Adult men and women in several occupational groups were given a series of tests covering educational ability (or "intelligence"), clerical aptitude, motor dexterity, and mechanical aptitude. Standard-score and percentile equivalents for each test were determined on a representative sampling of the em-

ployed population in three principal cities of Minnesota. The two contrasting profiles reproduced in Figure 79 were plotted from the average scores of a group of male office clerks and a group of garage mechanics.[3] It should be noted that these profiles have been plotted in terms of converted standard scores, with a mean at 5.0 points, as shown along the top of each graph. Along the base are the percentile values which correspond to these standard scores.[4] Figure 80 shows the average profiles of women office clerks and retail saleswomen.[5] Here, too, the contrast between the two groups is sharp.

That these patterns represent typical and consistent differences is indicated by several additional observations made in the same survey. In one check, the accuracy with which individual profiles could be assigned to the proper occupational category solely on the basis of the test scores was determined. With sample profiles of the corresponding occupational groups as standards of reference, the profiles of 158 employed women were examined. In this group, the profiles of 90 women office clerks and 68 retail saleswomen were mixed in random order. Correct identification of the occupational category was made in 92.4% of the cases.[6] Comparisons of occupational sub-groups varying in degree of job efficiency, as well as comparisons of groups employed by different companies, also indicated considerable consistency in the occupational profile.

The practical application of occupational ability patterns in vocational counseling is illustrated by the General Aptitude Test Battery of the United States Employment Service (12). The battery consists of 15 tests measuring 10 selected aptitudes, including general intelligence, verbal, numerical, and spatial abilities, clerical perception, form perception, and several motor tests. Through the testing of many groups of employed persons, occupational ability patterns have been established for twenty fields of work, representing approximately 2000 specific occupations. These ability patterns are described in terms of minimum scores in those aptitudes found to be significant for

[3] The number of garage mechanics was 102, although the clerical aptitude average is based on only 101 cases. Among the office clerks, the number taking each test ranged from 66 to 114.

[4] This correspondence is based on the assumption that the abilities in question are normally distributed.

[5] For the women office clerks, N varied from 164 to 180; for the saleswomen, from 65 to 137.

[6] The proportion of correct identifications to be expected by chance would be 50%.

each type of occupation. All the aptitude measures are expressed as standard scores with a mean of 100 and an SD of 20. Each individual to be counseled is given the entire battery, and his scores are compared with the patterns required for the different occupations. For example, accounting and related occupations were found to require a

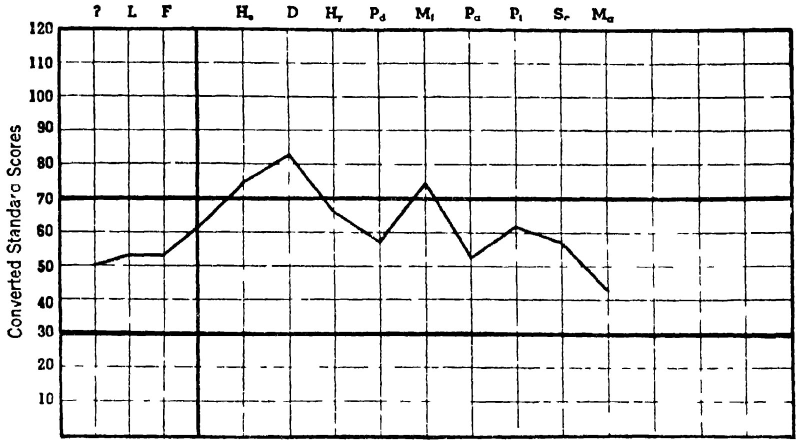


Fig. 81. A Personality Profile: Standard Scores of a Young Man of Eighteen on the Minnesota Multiphasic Personality Inventory. (Unpubl. data by courtesy of Dr. William C. Bier.)

minimum score of 130 in general intelligence and in numerical ability (pattern GN). Occupations grouped under heavy metal structural work, plumbing and related work, and wood structural work called for an NSM pattern, with a minimum score of 85 in numerical and spatial aptitude and in manual dexterity. Similar patterns have been worked out for twenty fields of work studied to date with this battery.

An increasing use of the psychographic approach is also evident in the field of personality testing. In place of a single index of "emotional adjustment" or instability, a multi-dimensional description of the individual in terms of several categories is substituted. An illustration of such a technique will be found in Figure 81, which shows the profile of a young man of 18 on the Minnesota Multiphasic Personality Inventory (18). This profile indicates a pronounced tendency toward depression (D), together with somewhat undue concern with health and physical complaints (Hs) and a deviation of the basic

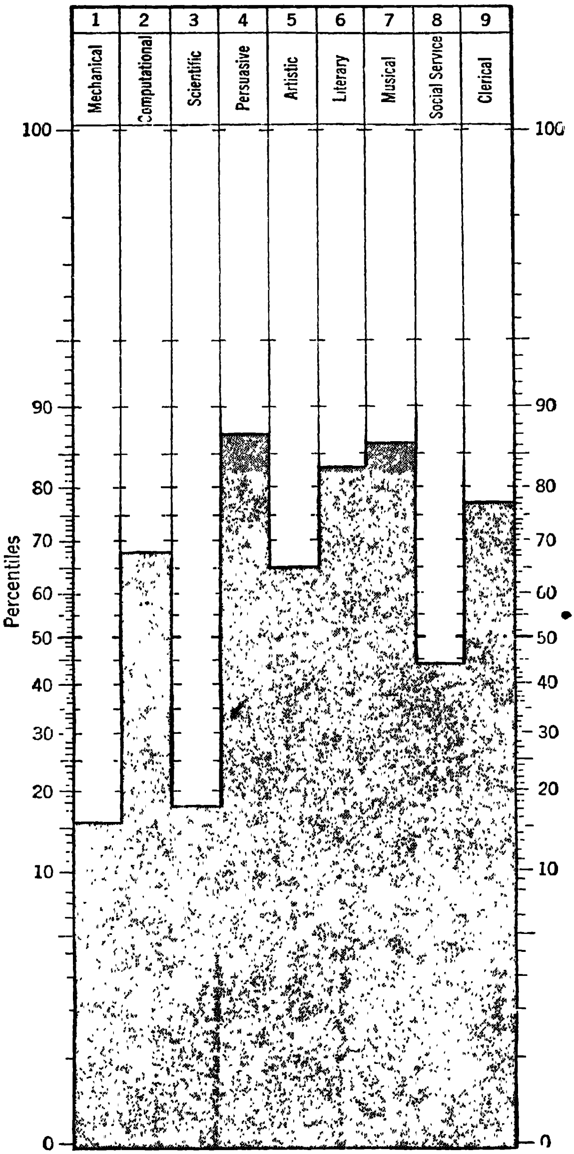


Fig. 82. An Interest Profile: Percentile Scores of a Male College Student on the Kuder Preference Record. (Unpubl. data by courtesy of Industrial Division, The Psychological Corporation.)

interest pattern in the direction of femininity (Mf). All the scores in this profile are expressed as standard scores, with the mean set at 50 and one standard deviation equal to 10 points. Thus, for example, a score of 75 falls 2.5 SD's above the average of the standardization sample.

Among the first tests to employ the profile method of scoring was the Allport-Vernon Study of Values (1). Mention should also be made of the various interest tests, such as those devised by Strong (45) and Kuder (25), whose approach is essentially psychographic. An illustrative profile depicting the Kuder Preference Record scores made by a young college man is presented in Figure 82. In this test, the individual indicates his like or dislike for a variety of activities, and his choices are subsequently classified into nine areas of interest, as shown on the chart. The subject in the present illustration shows a strong interest in persuasive activities (found to rate high among salesmen), as well as in literary and musical activities; his interests in the clerical, computational, and artistic areas are also quite high. He falls slightly below average in social service interest and far below in mechanical and scientific interests. It should be noted that in the Kuder profile chart the percentiles have also been spaced in conformity with the assumption of a normal distribution, as was done in the Differential Aptitude Tests discussed above.

EXTREME ASYMMETRIES OF TALENT

A consideration of individual cases which display conspicuous asymmetry of abilities helps us to visualize the extremes of variation which may occur within the individual. The study of special talents and defects may be regarded as one approach to the analysis of abilities and their mutual interrelationships. Are deficiencies along certain lines consistent with intellectual superiority? Do special talents in particular fields ever accompany general intellectual backwardness? The occurrence of special talents or defects in a given area of behavior would suggest that ability in that area may develop and vary independently of ability in other areas. These case reports serve to point up and vivify the findings of the more systematic statistical studies of trait relationships.

Musical aptitude tests seem to have little or no relationship to superior intelligence. This is illustrated by data obtained by L. S. Holling-

worth (22) in a study of 49 intellectually gifted children. All the subjects were enrolled in special classes conducted for children with IQ's of 135 or higher. The median Stanford-Binet IQ of this group was 153, and the range extended from 135 to 190. The children were tested with the Seashore Measures of Musical Talent. Their scores were evaluated in terms of fifth grade norms, since the chronological ages of the gifted group corresponded closely to those of the fifth grade school children in the Seashore standardization sample. Below will be found the average percentile score obtained by the gifted group on each of the Seashore tests.

Pitch	47
Intensity	50
Time	58
Consonance	48
Tonal memory	52

A percentile score of 50, it will be recalled corresponds to the middlemost score of the standardization group and thus represents a "normal," or average, performance. The fact that all the average percentile scores of the superior group were so close to 50 indicates that musical aptitude is distributed among these children in very much the same fashion as in any group of the same age chosen at random. Although in intelligence test performance these subjects were all within the upper 1% of the general population, their individual percentile scores on the music tests ranged from zero to 98.

Individual cases of intellectually superior children with a pronounced deficiency in music can readily be found. One 10-year-old school boy, for example, had an IQ of 151 but obtained scores which ranged from the zero to the 30th percentile on the Seashore music tests (20, p. 179). His school work in such subjects as reading, arithmetic, and elementary science was excellent. But his music teacher regarded him as a complete failure and advocated that he repeat the grade!

Case studies of arithmetical prodigies and "lightning calculators" indicate that a high level of numerical aptitude can likewise occur in individuals of average or inferior intelligence. Many such cases, from the early Greeks to the end of the last century, were collected and described by Scripture (40) and later by Mitchell (30). More recently, much of this material was brought together in a collection of papers prepared and edited by Bryan, Lindley, and Harter (9). In

regard to their achievements along other lines, or their practical ability to succeed in everyday life, mathematical prodigies run the gamut from genius and eminence of the highest order to mental dullness. A few would no doubt be classified as "borderline" or lower on current intelligence tests. At the other extreme are such men as Gauss and Ampère, whose exceptional talents covered a wide range, and who made distinguished contributions in mathematics and allied fields. These men were "lightning calculators," but also possessed very superior ability along many other lines. For the present purpose, however, we are concerned with cases of asymmetrical development in which prodigious arithmetic powers are coupled with mediocrity or deficiency in other respects.

Henri Mondeaux (cf. 40), the untutored son of a poor woodcutter, is a famous example of remarkable arithmetic ability in an otherwise dull person. In his childhood he received no instruction, but was sent to tend sheep at the age of 12. While engaged in this occupation, he amused himself by counting and arranging pebbles; by this means he learned to carry out arithmetic operations. He worked out for his own use many special devices and aids to computation. After long exercises at these calculations, he offered to tell people he met the number of seconds in their ages. At this time, a schoolmaster became interested in him and offered to instruct him. Unfortunately the boy had a very poor memory for names and addresses and he spent nearly a month searching the city before he was able to locate his benefactor. Mondeaux was subsequently exhibited by his teacher at several colleges and universities and in 1840 was presented before the Academy of Sciences at Paris. His was not merely a talent for routine calculation, but he demonstrated his ability to solve, by ingenious devices of his own making, complex problems such as the following:

> There is a fountain containing an unknown quantity of water; around it stand people with vessels capable of containing a certain unknown quantity. They draw at the following rate: the first takes 100 quarts and 1/13 of the remainder; the second takes 200 quarts and 1/13 of the remainder; the third, 300 quarts and 1/13 of the remainder, and so on until the fountain is emptied. How many quarts were there?

Mondeaux gave the correct answer to this problem in a few seconds and then explained the method whereby he had arrived at the solution.

A similar case is that of Tom Fuller, an African slave brought to

America in 1724 at the age of 14. He could neither read nor write and received no formal instruction. As in the case of Mondeaux, his arithmetic was entirely self-taught. It is reported of him that when asked how many seconds a man had lived who is 70 years-17 days-12 hours old, he gave the answer, after 1½ minutes, as 2,210,500,800 seconds. One of his questioners had meantime been computing with paper and pencil and reported that he had arrived at a different number, which he proceeded to read off. At this, Fuller immediately pointed out that his questioner had forgotten to allow for leap years!

A few cases of "lightning calculators" have been directly observed and investigated by psychologists.[7] The information thus obtained, as well as the careful analysis of available reports on arithmetic prodigies, has brought to light certain characteristics of these individuals which may account for their talent. In most cases, the individual has worked out a number of short cuts and special devices which enable him to compute far more efficiently than is ordinarily possible. Secondly, such individuals have usually memorized many more number combinations, such as squares, cubes, roots, and products, than are at the disposal of the average man. Arithmetical prodigies invariably manifest a very keen interest and fascination for numbers. As a result, they devote much time to analysis of computation methods and to drill which would otherwise prove highly monotonous. Many also seem to have a large perception span which enables them to grasp a long series of numbers simultaneously, as well as vivid imagery, making possible "mental computation" without the aid of paper and pencil. Moreover, such prodigies often build up a wealth of associations to numbers, and are thus able to use numbers in their thinking in much the same way that the average person uses verbal symbols.

The most spectacular examples of special talent are the so-called *idiots savants*. This term, which literally means "wise idiots," is somewhat misleading, since the usual idiot savant is neither particularly wise nor an idiot. He is not sufficiently deficient to be classified as an idiot, but is frequently found at the moron or borderline level. And he is "wise" only in a very limited field. In the practical management of his own life he is ordinarily a complete failure.

As might be expected, idiots savants are extremely rare. Because of their remarkable accomplishments, however, they attract considerable attention, and a number of fairly complete descriptive accounts are

[7] Cf. 5, 9, 26.

now available. Most of the earlier cases were summarized by Tredgold (50). More recently, Rife and Snyder (37) addressed an inquiry to 55 American institutions for the feebleminded, through which they were able to locate 33 cases of idiots savants. Of these, 8 showed a special talent in music, 8 in mathematics, 7 in drawing, and 10 in miscellaneous areas including mechanics, memory, and motor coordination. A few of these cases manifested skills which were narrowly limited and of dubious psychological significance. On the other hand, a number gave evidence of well-rounded achievement in a fairly broad area.

Several cases of special talent in pictorial art have been found among the feebleminded. Such individuals are able to execute excellent reproductions of well-known paintings. Occasionally this talent passes beyond mere copying and suggests real creative genius. Such a case is that of Gottfried Mind (50, Ch. XV), diagnosed as a cretin imbecile. His mental deficiency, manifested from an early age, was such that he was unable to learn to read or write. His movements were awkward, his hands large and rough, and his general appearance that of the traditional mental defective. Since he showed considerable talent for drawing, he was given some instruction in this field. His subsequent success in pictorial art was phenomenal. Because of his excellent drawings of cats, he came to be known as "The Cat's Raphael." In addition, he produced drawings and water-color sketches of deer, rabbits, bears, and groups of children, which were remarkable for their life-like quality and masterly execution. His fame spread throughout Europe and one of his pictures of a cat and kittens was purchased by King George IV of England.

An equally remarkable case is that of J. H. Pullen, who has been called "The Genius of Earlswood Asylum" (50, Ch. XV). This individual had extraordinary mechanical ingenuity coupled with talent in drawing and carving. In other respects he was very deficient. He did not talk until the age of 7, and for a long time uttered only the word "muvver." Probably because of a severe hearing deficiency, his speech never came up to normal. He was taught by his family to write and to spell the names of simple objects, and this constituted the extent of his schooling. From an early age, he spent much of his time in drawing or in carving ships out of pieces of firewood, occupations in which he showed considerable proficiency. At the age of 15, he was admitted to Earlswood Asylum, where he was put to work in the carpenter's

shop and soon became an expert craftsman. During his sixty-six years at the asylum, he produced an impressive array of beautiful and highly ingenious objects, including crayon drawings, carvings in ivory and wood, excellent models of ships, and various mechanical devices. Occasionally he even designed his own instruments to help him in his work.

One of Pullen's constructions was a representation of a gigantic human form, thirteen feet high. This full-fashioned "robot" could be made to execute a variety of movements, such as raising the arms, rotating the head, protruding the tongue, and opening and shutting the mouth or eyes. Another remarkable construction was a model of a ship, beautifully executed in the minutest detail. This model required over three years for its completion and attracted universal admiration when exhibited. Pullen's work revealed artistic imagination as well as mechanical ingenuity, skill in planning, and painstaking execution. Being cut off from many ordinary sources of stimulation by deafness, it is probable that he concentrated all his efforts from childhood upon the development of this one remarkable talent. In regard to general personality development, he is described as childish and immature, emotionally unstable, and lacking in common sense.

Special talent in music has also been observed among the intellectually deficient. A case (50, Ch. XV) of exceptional musical ability combined with serious defect in other respects is that of a woman in the Salpêtriere, a famous French institution for the feebleminded and the insane. This patient was an imbecile, blind from birth, a cripple, and affected with rickets. She was, however, able to sing without error any selection which she had heard. It became customary for her fellow-inmates to come to her so that she might correct their mistakes in singing. She attracted wide attention, and it is reported that the composers Liszt and Meyerbeer visited her "singing class" to bring their encouragement and consolation.

More recently, another instance of musical talent combined with intellectual deficiency was investigated by the use of standardized intelligence tests (cf. 29). This was the case of a boy admitted to a feebleminded institution at the age of 14. He came from an intellectually superior family which included many musically gifted individuals among its members. As a child, the subject was intellectually normal and manifested his musical talent from an early age. When three years old, he had pneumonia and meningit. , and since that time

he underwent steady mental deterioration. Upon admission, his IQ was 62; at the last testing, it had dropped to 46. He was then over 20 years of age and had a mental age of 7 years-5 months. His memory was unimpaired, however, and he retained his excellent musical ability. Although never known to compose a piece, he could play difficult music by ear and was also able to read difficult musical compositions at sight.

The feats of memory performed by some feebleminded individuals have often attracted notice. Tredgold (50, Ch. XV), for example, describes a 65-year-old mental defective in Earlswood Asylum with a remarkable memory for historical facts. He could repeat the dates of birth and death and the essential facts in the life of any prominent character in history. This knowledge was acquired largely by rote, through poring over all available books on biography and history. It was not, however, a matter of sheer meaningless repetition, as was shown by the subject's responses when questioned on the material. Another patient at the same institution showed an excellent memory for dates and occurrences which had come within his own experience. He proved a useful source of reference on local happenings in the institution.

Arithmetical prodigies have also been found among the ranks of the feebleminded. Usually the skill manifested is confined to the mechanics of computation. Thus the subject may perform long and complicated calculations within a very short time and without the aid of paper and pencil. A favorite feat is to determine the number of minutes a person has lived, from a knowledge of his age or date of birth. Multiplication of three-place numbers, naming square roots and cube roots of four-place numbers, and similar difficult operations have also been executed within a few seconds. In some cases, this numerical aptitude goes beyond routine computation, as is indicated by the individual's ability to solve mathematical problems expressed in fairly elaborate and confusing terms. An example of fairly complex numerical aptitude, appearing early in life, is provided by the case of a 27-year-old man with a mental age of 3, described by Rife and Snyder (37, pp. 553-554). They write:

> As a small child he would scribble figures on the bathroom tiles or other places whenever he could get hold of a pencil. He never learned to talk, and even now cannot perform such simple requests as pointing to his eyes or ears. In school he could do absolutely nothing, so was sent

home, and at sixteen was admitted to the Institution. His hearing is normal. . . . Although he is incapable of carrying on a conversation, or of understanding spoken requests, one may make one's desires along mathematical lines known with a pencil. When a pencil and paper were taken, and the figures 2, 4, and 8 written in a vertical column, the patient immediately continued the series 16, 32, 64, etc. When the series 2, 4, 16 was started, he immediately continued this one, the sixth number being 4,294,967,296. Then 9—3 was written, in the attempt to indicate square root. Under this, several numbers such as 625, 729, and 900 were written. The square root of each was immediately and correctly written. Any problem of multiplication of several digits by several digits was done immediately, only the answer being written.

THE MEASUREMENT OF TRAIT VARIABILITY

The term *trait variability* was first proposed by Hull (23) to designate variability from trait to trait within the individual, in contrast to *individual variability*, which refers to the differences among individuals in a single trait. Any of the methods commonly employed to measure individual variability can be applied to the measurement of trait variability, provided that the scores on different tests are expressed in the same units. In view of the extreme asymmetries of talent occasionally observed, the question arises regarding the *extent* of trait variability found among people in general.

In an early study by Hull (23), the extent of trait variability was gauged by comparing it with the amount of individual variability within single tests. The data consisted of the scores of 107 high school freshmen on 35 tests, including several sub-tests from intelligence scales, as well as tests of motor functions, attention, perception, and personality characteristics. All scores were transmuted into standard scores and the SD of each person's 35 scores was then computed. There were thus obtained as many SD's as there were subjects, viz., 107. The average of these 107 SD's was 6.33. Hull compared this figure with the SD of 7 which, in the scale of units employed, represented the variability from person to person (i.e., individual variability) in any one of the tests. After allowing for possible chance errors of sampling and measurement, Hull estimated that the trait variability was about 80% as great as the individual variability.

It should be noted that such an estimate of the extent of trait varia-

bility is limited by the conditions of the particular investigation. The estimate will be affected by the number of tests employed, the range of functions tested, and the heterogeneity of the group (16). The more homogeneous the group, the smaller will be the SD's representing individual variability in any of the tests. It will be recalled that these SD's provide the units in terms of which the individual's score in each test must be expressed before his trait variability can be measured. It follows that the more heterogeneous the group, the smaller will the trait variability appear, since it will be expressed in terms of a larger unit. Thus such estimates should not be generalized beyond the type of population on which they were obtained.

That the 80% estimate is not too far from what would be found in other typical groups is suggested by the findings of a more recent study conducted in France (34). The average trait variability was measured in each of four different groups, the number of tests used in each group being shown below:

Subjects	*N*	*Number of Tests*
Flight candidates	148	25
Vocational school girls	171	15
Apprentices	1274	8
Paris school children	693	5

From the results obtained with these groups, it was concluded that trait variability tended to be slightly over 75% as great as the individual variability of the group.

The distribution of each individual's scores on the different tests seems to follow the general form of the normal curve. Most of the individual's scores cluster about his own average, with only a few scores deviating widely in either direction (23). The extent of trait variability differs widely from person to person, some individuals being considerably more uniform in their abilities than others. In the Hull (23) study, for example, the individual SD's for trait variability ranged from 4.30 to 9.09. These indices of trait variability are themselves normally distributed, and show no evidence for the presence of distinct "types" in reference to degree of trait heterogeneity (34).

A question of considerable theoretical as well as practical interest is whether any relationship exists between *ability level* and extent of trait variability in different individuals. Some investigators have found

no significant correlation between the two (7, 23). In these studies there appeared to be no relationship between the individual's trait variability and his standing either in specific tests or in the average of all the tests. The groups studied, however, consisted of either high school students or college freshmen and were therefore relatively homogeneous in ability. It is possible that the correlation might be higher if a wider ability range were considered. Moreover, the relationship between ability level and trait variability may be curvilinear,[8] in which case the use of the usual Pearson correlation coefficient would underestimate the degree of relationship.

There is some evidence in the data collected by other investigators which suggests a negative correlation between ability level and trait variability (34, 49). In one study (49), the duller children showed more unevenness of abilities than did the average, and the average were more uneven than the bright. When the trait variability of the average group was taken as 100%, the trait variability of the dull group was found to be 110% and that of the bright group 92%. Although the trend toward an inverse relationship between intelligence and scatter was found consistently with all the intelligence tests employed in this study, the actual correlation was low. Similarly, in the French study on four different groups cited above (34), a consistent tendency was observed for trait variability to be greater among the poorer subjects. Below are the average trait variabilities of subjects in each of the four groups, the subjects being classified according to their composite performance level on all the tests.[9]

Total Performance on All Tests	*Flight Candidates*	*Vocational School Girls*	*Apprentices*	*Primary School Children*
Superior (Upper 25%)	2.83	3.43	3.15	1.23
Average (Middle 50%)	3.4	3.66	3.19	1.64
Inferior (Lower 25%)	4.56	3.83	3.42	1.67

[8] By curvilinear relationship is meant that the type of relationship is not uniform throughout the range. For example, salaries may increase as amount of education increases up to a point and then decrease as the highest educational levels are reached. Or the *rate* of increase may vary in different parts of the range. In Bowns's study of trait variability (7) curvilinear correlation was computed, but still no evidence of significant relationship was found. In this study, however, the subjects were quite homogeneous, all being college freshmen.

[9] The unit in which these values are reported is ¼SD. Since most of these trait variabilities are close to 3, it will be noted that they are approximately ¾ as large as the SD of the distribution of individual differences. The relatively low trait variabilities found among the school children are probably the result of the small number and restricted variety of tests employed with this group.

Relevant data are also provided in the study conducted by DeVoss (10) to determine whether gifted children are more specialized in their abilities than normal children, i.e., whether they show wider trait variability. A group of 100 subjects were selected on the basis of mental age from a larger group of "gifted children" studied by Terman and his associates (cf. Ch. 17). The mental ages of DeVoss' group ranged between 14 and 15-5, with an average of 14-8; the chronological ages ranged from 8–6 to 11–1. The average IQ was 149.4, and the range from 136 to 180. In school grade, the children were scattered from the third to the eighth grade, inclusive. These subjects were compared with a control group of 96 unselected eighth grade school children of approximately the same *mental ages* as the superior group. Both groups were given the Stanford Achievement Test, consisting of seven sub-tests on different school subjects, as well as information tests in special fields. All scores were reduced to standard measures.

Examination of the inter-test variations within each subject's scores revealed many differences large enough to be significant. By means of a specially devised statistical formula,[10] it was possible to estimate how large a trait difference might be obtained simply through the operation of chance factors, such as inadequacies in the tests employed. Upon the application of this formula, it was discovered that a large percentage of the trait differences fell beyond the chance limits and must therefore represent a true discrepancy in the individual's standing in the traits compared. In Table 27 are given the percentage of trait differences, in both the gifted and control groups, which fell outside of the distribution of differences expected by chance. The percentages in the gifted group are given above the diagonal, those in the control group below the diagonal. The tests which are being compared are indicated in the top row and first column to the left. Thus, in the gifted group, 24% of the differences between scores on arithmetic reasoning and computation fell beyond the chance limits; the corresponding percentage for the control group is 34%, and so on.

[10] A formula for the computation of the PE of the difference between an individual's standard scores on any two tests:

$$PE_{diff} = .6745\sqrt{2 - r_{I\,I} - r_{II\,II}}\ ,$$

in which $r_{I\,I}$ and $r_{II\,II}$ are the reliability coefficients of the two tests.

TABLE 27 *Percentage of Trait Differences among 100 Gifted and 96 Control Children Which Fall Outside of the Chance Limits* *

(From DeVoss, 10, p. 325)

Tests to Be Compared	*Arithmetic computation*	*Arithmetic reasoning*	*Word meaning*	*Sentence meaning*	*Paragraph meaning*	*Language usage*	*Spelling*	*Science information*
Arithmetic computation		24	32	31	34	37	26	33
Arithmetic reasoning	34		30	31	31	38	29	29
Word meaning	39	30		13	24	25	26	23
Sentence meaning	40	26	13		26	24	28	29
Paragraph meaning	35	26	14	17		33	27	34
Language usage	39	33	24	23	17		30	32
Spelling	31	36	28	33	25	30		30
Science information	35	20	24	25	21	30	31	

* Gifted group above the diagonal: upper right-hand block. Control group below the diagonal. lower left hand block.

It will be noted that in every test pair compared there are found differences over and above those expected by chance. This is true of both gifted and control groups. The percentages of such differences are also closely similar in the two groups, test by test. In the gifted group, these percentages vary from 13 to 38, in the control group from 13 to 40. The average percentages are 28.89 and 27.82 for gifted and control groups, respectively. Out of the 28 inter-test comparisons given in Table 27, the gifted group has the larger percentage of excess differences 13 times, the controi group has the larger percentage 12 times, and the two have identical percentages in 3 cases. Thus there seems to be no appreciable or consistent difference in trait variability between intellectually normal and superior children.

In conclusion, the available data on trait variability offer no support to the popular notion that intellectually gifted children show a higher degree of specialization than do the normal. In fact, if any

difference exists in this respect, it is the duller individual who appears to be more specialized than the normal, but additional research is needed to establish such a relationship.

Two additional variables whose relationship to trait variability has been studied are *practice* and *age*. In a re-analysis of data collected by several investigators, Preston (36) has shown that trait variability tends to decrease with practice and to increase with age. The effect of equal practice is to make the subject more uniform in the various practiced tasks. Age has the opposite effect upon trait variability, the older individual showing more scatter or specialization of ability. It cannot be assumed, of course, that age *per se* accounts for such changes in trait variability. The groups compared in these investigations also differed in educational level and probably in other respects. It is entirely possible, for example, that education may increase trait variability, even though practice tends to decrease such variability. Education obviously does not consist of "equal practice" in all intellectual functions. Not only does the amount of practice vary in different areas, but motivational changes and other complicating influences are probably introduced in different ways for different individuals. The effects of education on trait variability may thus be quite unlike those obtained in simple practice experiments.[11]

The relationship between *personality characteristics* and unevenness of ability also presents a fruitful area for research, but the available data on this question are still highly tentative.[12] Interest in this relationship has been stimulated in recent years by the development of the Wechsler-Bellevue Intelligence Scale. It has been suggested that the pattern of the individual's scores on the different sub-tests of this scale may serve as an index of various emotional disorders. This application of the concept of trait variability has attracted wide attention among clinicians, but it would be premature to evaluate it in its present stage.

INTERCORRELATIONS OF TEST SCORES

The examination of extreme examples of asymmetrical development, as well as the measurement of trait variability within individuals in

[11] The relationship of age and education to the specialization of ability will be considered more fully in the following chapter.

[12] Cf., e.g., Bown (7), Freeman (14).

general, suggests that superior talents in one line may be associated with inferior abilities in other respects. It is not to be concluded from this, however, that compensation is the rule. Superior standing in one trait does not *imply* inferiority in another. We have cited only examples in which individuals with a high standing in a certain trait A make a poor showing in a second trait B. We could with equal facility find cases in which the individual is superior in A as well as B, or superior in A and average in B. This, in fact, is what we mean by a *low or zero correlation*. If various abilities are specific and mutually independent, so that an individual's standing in one tells us nothing about his relative standing in another, we should expect the correlation between such abilities to be zero or very low.

Correlation thus offers another approach to the analysis of trait variability. It should be noted that these are literally alternative approaches or ways of expressing the *same facts*. Thus the asymmetries of ability illustrated in an earlier section are only extreme cases of trait variability. Similarly, it can be shown by simple algebra that measures of trait variability depend upon the intercorrelations among the traits under consideration, and that the one type of measure can be derived from the other (cf. 16, 35). The average trait variability of a group of individuals in a given series of tests can be found by the following formula:

$$\bar{V} = 1 - \frac{1}{n} - \frac{\Sigma r}{n^2}$$

in which,

- $\bar{V}$ is the average variance[13] within the individual, expressed in terms of the variance among individuals in a single test,
- n is the number of tests, and
- Σr is the sum of all the intercorrelations among these tests, each correlation being entered twice. For example, the correlation between tests 1 and 2 would appear in the complete correlation table as as r_{12} and as r_{21}.

By means of this formula, it can readily be shown that if all the tests were perfectly correlated with each other, trait variability would be zero (35). On the other hand, if all the intercorrelations among the tests were zero, the average trait variability would be nearly as high as the individual variability, and would approach the latter as the

[13] The variance is the square of the standard deviation.

number of tests increases. An examination of correlation coefficients can thus provide the same type of information which is obtained by the measurement of trait variability.

Profile asymmetries can likewise be investigated by the correlation technique. Between what areas of ability are such asymmetries likely to occur? Do certain functions tend to vary together within the individual, so that a deviation in one will be accompanied by a similar deviation in the other? These are the types of questions that are answered by correlation coefficients. Certain functions have long been recognized as "special aptitudes," a designation which carries with it a tacit presupposition of low or zero correlation with other functions. Among the most familiar are musical, artistic, and mechanical aptitudes. It will be recalled that these are some of the areas in which marked asymmetries of talent have been reported. We may now inquire what the correlational approach has to offer regarding these aptitudes.

Tests of *musical aptitude* have consistently shown low—and usually negligible—correlations with measures of "general intelligence." In a group of 74 college students, a correlation of −.17 was found between intelligence test scores and a test of musical appreciation (19). In the same group, a different form of the music test yielded a correlation of −.15 with the intelligence test. The fact that these correlations are negative might suggest a slight tendency for the more "intelligent" individuals within such a group to be poorer in music appreciation, but the correlations are too low to be significant. In another investigation with 230 college students, intelligence test scores were correlated with each of the Seashore tests of musical sensitivity. The correlations were all positive but very low, only one being significantly higher than zero (13). Other studies with the Seashore tests have yielded equally low correlations with intelligence tests at all age and grade levels (38; 31, pp. 335–340).

Aptitude in *pictorial art* shows a similar independence of general intelligence. Correlations ranging from −.14 to +.28 were found between the Meier Art Judgment Test and intelligence test scores in various groups of high school and college students (28). None of these correlations was statistically significant. Among elementary school children, an equally low relationship has been found between intelligence test scores and test performance in either representative (6) or

creative drawing (47). A few of these correlations are statistically significant, indicating more than a chance relationship, but all are low enough to permit marked asymmetry in individual cases.

Mechanical aptitude also appears to be a special ability. The Stenquist Assembly Tests, involving the construction of common mechanical objects such as lock, bicycle bell, and mouse trap, correlated .23 with intelligence test scores in a group of 267 seventh and eighth grade boys (43). Although significantly higher than zero, this correlation indicates only a slight degree of relationship. In the standardization of the Minnesota Mechanical Ability Tests (32), a correlation of .13 was found in a group of 100 junior high school boys between IQ and a mechanical aptitude battery consisting entirely of apparatus or manipulation tests. Similarly, an investigation on 225 college men showed a correlation of only .07 between a vocabulary test and the Minnesota Paper Form Board Test (2). The latter is a paper-and-pencil test of the ability to visualize spatial relations. Vocabulary tests, which measure the subject's understanding of word meanings, have been found to correlate so highly with the majority of common intelligence tests as to be practically interchangeable with them. From these examples, it is apparent that in large groups of subjects of different age and academic level, only a very low positive correlation exists between spatial or mechanical ability and the verbal type of intelligence test. When the mechanical problems are presented verbally, the correlations with intelligence tests are usually higher because of the common influence of the comprehension of verbal directions, knowledge of words, and general facility with verbal material.

It should be noted that superior *achievement* in the fields of art, music, or mechanical work requires other abilities in addition to the special talents which have been discussed. Many of the abilities measured by intelligence tests are needed for the type of training which makes higher levels of accomplishment in these fields possible. The professional application of these talents, moreover, often demands other skills in the comprehension of complex verbal or numerical concepts, effective social communication, and the like. It is not surprising, therefore, that when indices of actual achievement are employed, successful performance in music, art, or mechanical work shows a closer relationship with intelligence test scores than is found between *tests* of these special aptitudes and intelligence tests.

Among school children, fairly high correlations have been reported

between intelligence test scores and such measures of musical achievement as grades in music classes or ratings by music teachers (31, pp. 335–340). That these correlations are higher than those found with musical aptitude tests may result in part from the contribution of other abilities besides musical aptitude in determining such achievement and in part from a probable halo effect in the grades and ratings. For successful accomplishment in almost any specialized area, a certain minimum level of "general intelligence" is an essential prerequisite. In surveys of artistically talented high school students as well as recognized adult artists, the average intelligence test performance was found to be clearly superior to that of comparable, artistically unselected samplings (48). Similarly, a group of artistically superior children studied at the University of Iowa had IQ's ranging from 111 to 166 (27). Within such groups, however, there is no evidence that the degree of artistic merit or recognition is correlated with intelligence test score. Similar data for the mechanical field are provided by the average intelligence test scores obtained by persons engaged in different vocations. Inventors, engineers, architects, and other persons occupied with creative mechanical pursuits receive considerably higher intelligence test scores than the average.

All this simply suggests, however, that more than a single specialized talent is required for the higher levels of achievement in music, art, or mechanics. The fact remains that a highly developed ability in any of these areas may coexist with low ability along other lines. In such cases, the specialized ability simply does not have as much "social market value," either academically or vocationally, as it would if it were accompanied by other abilities.

Of the principal aptitudes suggested by the case reports of extreme asymmetries of intellectual development, only *numerical ability* remains to be examined. Despite the indisputable presence of "mathematical prodigies" who are deficient in other respects, numerical ability has not usually been classed with special aptitudes. Arithmetic tests are also frequently included in intelligence scales. Recent correlational analysis has demonstrated, however, that the relationships *between* verbal and numerical tests are much lower than those *within* either group of abilities. In many investigations, the correlations between verbal and numerical tests were no higher than those between verbal tests and the various special aptitudes discussed above.

In one study (39), 210 college men were given five verbal and

four numerical tests. The average correlation between all possible pairs of verbal tests was .49; the corresponding average correlation for the numerical tests was .34. When verbal and numerical tests were paired off, the average of the correlations thus obtained was only .14. Even this rather low correlation probably resulted in part from the fact that in at least one of the numerical tests the problems were expressed in verbal terms, and in all the tests the directions were given verbally. Among 225 college men tested in another investigation, a correlation of −.01 was found between arithmetic reasoning and vocabulary (2). In still another study (3), 140 college women were tested with two verbal tests (vocabulary and analogies) and two numerical tests (arithmetic reasoning and number series completion). The correlation between the two verbal tests was .65 and that between the two numerical tests .58. When verbal and numerical tests were paired off, however, the average correlation was only .16.

These various findings suggest that verbal and numerical tests appear to be measuring two "special aptitudes" in the same sense as the other tests discussed above. To be sure, these two abilities are less specialized among younger subjects and individuals of lower educational level, a finding whose implications will be considered in the following chapter. But among all individuals, they are sufficiently differentiated to have led to the now common practice of reporting separate scores for "linguistic" and "numerical," or "quantitative," intelligence on most intelligence tests.

WHAT DO "INTELLIGENCE TESTS" MEASURE?

After this brief overview of the most commonly observed "special aptitudes," we may well ask what is left of "general intelligence." Perhaps the content of "intelligence tests" may provide a clue to the connotations of the term "intelligence" as a working concept. It will be recalled that the original aim of intelligence tests was to sample a large number of different abilities in order to arrive at an estimate of the subject's general level of performance. In so far as the individual's standing in specific functions differs, such a general estimate is unsatisfactory. It is apparent, however, that current intelligence tests do not even furnish an adequate estimate of the average ability of the individual, since they are overweighted with certain functions and omit others. Thus in the non-language and performance tests of

intelligence, spatial aptitude plays the dominant role. Most paper-and-pencil tests, on the other hand, measure chiefly verbal ability and, to a slighter extent, numerical ability. Since the latter type of test is by far the most frequently employed, the term "intelligence" has come to be used almost synonymously with verbal ability. Mental age on the Stanford-Binet, for example, has been found to correlate on the average about .81 with performance on the vocabulary test of the scale (46, p. 302). Within single age groups, this correlation ranges from .65 to .91.

From another angle, most intelligence tests may be regarded as measures of scholastic aptitude, or *ability to succeed in our schools*. This is illustrated particularly well by the procedure commonly followed in validating intelligence tests. The term "validity," it will be recalled, denotes the degree to which a test actually measures what it purports to measure. In the case of most intelligence tests, validity has been checked against school success as a criterion. Scores on the test are correlated with school grades or teachers' estimates of ability, and the higher these correlations the more valid the test is said to be.

It should also be noted that tests of intelligence correlate nearly as highly with tests of school achievement as they do with each other. For example, the Stanford Achievement Test, a standardized examination in such school subjects as reading, arithmetic, spelling, history information, and the like, yielded a correlation of .66 with the National Intelligence Test, .71 with the Illinois General Intelligence Test, and .79 with the Otis Intermediate Test of Intelligence (24, Ch. 17). The correlations of different intelligence tests with each other run no higher than these, most of them falling between .60 and .80.

All in all, it is apparent that most intelligence tests are heavily weighted with certain functions, predominantly verbal aptitude. At the same time, they have proved to be of considerable value as empirical instruments of prediction in a wide variety of practical situations. In forecasting academic promise, aiding in the selection of applicants for jobs, and assisting the vocational counselor, they are making a significant contribution. Their usefulness in the large-scale problems of selection and classification of military personnel in World War II was indisputable. That such tests have proved to have empirical validity suggests that the criteria themselves may be "overloaded" with certain aptitudes. If the tests were not overweighted with verbal ability, their validity might drop appreciably, since verbal aptitude undoubtedly

plays a predominant role in determining successful achievement in our schools, our vocational pursuits, and other everyday life situations in our culture.

A CULTURAL CONCEPT OF INTELLIGENCE

Among the many definitions of intelligence which have been proposed by psychologists,[14] two concepts recur most frequently. First, intelligence is characterized as the ability to deal with abstract symbols and relationships. Secondly, it is described as the capacity to adapt to new situations or to profit by experience and is virtually identified with learning ability. Most of these definitions suffer from the weakness that, in their effort to be all-encompassing, they really tell us very little. If, for example, we define intelligence as the capacity for abstraction, we are immediately confronted with the fact that the same individual may deal effectively with abstract verbal concepts but be very deficient with quantitative concepts, or vice versa. Similarly, the available evidence offers no support for the view that "learning" is a unitary function (cf., e.g., 52). If intelligence were to be defined in terms of the ability to learn, a legitimate question would be, "To learn what?"

It is thus apparent that "intelligence" can be defined only with reference to a particular setting or environmental milieu. This viewpoint immediately suggests that there are, not one, but many definitions of intelligence. Within our cultural setting, intelligence apparently consists in large part of verbal ability. It will be recalled that the one field from which idiots savants are conspicuously absent is the linguistic one. Success in the practical business of everyday life—for both child and adult—is so closely linked with verbal aptitude that a serious deficiency in this respect will brand the individual as mentally incompetent. Conversely, the person who is especially proficient in verbal functions may thereby compensate for deficiencies along other lines and will rarely, if ever, find his way into an institution for the feebleminded. No other single talent seems to be such a saving grace in our society. Because of its intimate association with our concept of "general intelligence," verbal aptitude does not ordinarily enter into our classification of special talents or defects. Children who are defi-

[14] For a survey of different definitions of intelligence—early and recent—cf. 15, 17, 33, 41, 42, 44, 51.

cient in reading or verbal expression are usually inferior on intelligence tests (20). On the other hand, case studies of juvenile authors have invariably shown them to be children of very high IQ (21). To define intelligence within our culture is primarily to catalogue those activities which are made possible by linguistic development.

In summary, the data of the present chapter suggest that the individual's abilities in the verbal, numerical, spatial, musical, and artistic fields are relatively independent of each other. Of these, the verbal and—to a lesser extent—the numerical aptitudes are most closely identified with the concept of "intelligence" in our culture. In the chapter which follows, we shall inquire more intensively into the identification and interrelation of the psychological "traits" which have been suggested in the present chapter.

CHAPTER

15

Trait Organization

In its most elementary terms, a trait may be regarded as a category for the orderly description of the behavior of individuals. The concept of trait is concerned with the organization and interrelationships of behavior. Traits are therefore identified by observing or measuring varied behavior manifestations of the individual. Traits also refer, as a rule, to relatively enduring characteristics, which thus have some predictive value. Moreover, they usually cover those characteristics in which individuals differ appreciably from one another. Lastly, a cultural frame of reference is also evident, although not always stated, in most trait classifications. It is those aspects of behavior which are important within a particular culture or environmental setting which are generally identified and described as traits.[1]

Theories of trait organization are very old. As long as philosophers have discussed the nature of "mind," they have proposed theories regarding the units into which the "mind" was subdivided. With these speculations, however, we are not concerned. It is only since the application of psychological tests and quantitative methods that the relationships among the varied behavior manifestations of the individual could be measured. The more recent theories have been developed as interpretations of specific evidence and thus have a more empirical foundation.

MAJOR THEORIES

The Two-Factor Theory. The problem of trait organization was first placed upon an empirical basis with the publication of Spearman's 1904 article (88) in which were presented a theory and a new method of investigation. According to the original formulation of Spear-

[1] Even the *trait names* in our language have a cultural origin and in turn influence our selection and definition of traits.

man's Two-Factor theory (89, 91), all intellectual activities have in common one fundamental function which is called the *general* factor, or *g*. In addition, each activity has *specific*, or *s*, factors. The *s* factors are considered to be exceedingly numerous and strictly specific to each activity of the individual. No two activities can share specific factors, by definition. Spearman argued that such a theory is consistent with correlation results. Thus the presence of different specifics in every activity would explain the absence of perfect +1.00 correlations: no two activities, however much they may depend upon the *g* factor, are entirely free from specifics. The fact that most abilities are positively correlated, on the other hand, is attributed to the ubiquitous *g*. Different proportions of *g* and *s* in each activity would produce a wide range of positive correlations, all higher than zero and lower than 1.00.

It follows from the Two-Factor theory that the aim of mental testing should be to measure the amount of each individual's *g*. If this factor runs through all abilities, it furnishes the only basis for prediction of the subject's performance from one situation to another. It would be futile to measure specific factors, since each operates in only one activity. Accordingly, Spearman proposed that a single test, highly "saturated" with *g*, be substituted for the heterogeneous collection of items in intelligence scales. He suggested that tests dealing with abstract relations, such as the analogies test, are probably the best single measures of *g* and could therefore be employed for this purpose.

In regard to the nature of *g*, Spearman offered only tentative hypotheses. He proposed that *g* may be regarded as the general "mental energy" of the individual and the *s* factors as the "engines" through which it operates, or the specific neurone patterns involved in each activity. This interpretation of *g* and *s* is not, however, an integral nor a basic part of the Two-Factor theory. It might be noted that Spearman's *g* would also furnish a basis for the popular notion of general intelligence.

Even from the outset, Spearman realized that the Two-Factor theory must be qualified. When the activities compared are very similar, a certain degree of correlation may result over and above that attributable to the *g* factor. Thus in addition to the general and specific factors, there might be another, intermediate class of factors, not so universal as *g* nor so strictly specific as the *s* factors. Such a

factor, which is common to a group of activities but not to all, has been designated a *group factor*. In the early formulation of his theory, Spearman admitted the possibility of very narrow and negligibly small group factors. Following subsequent investigations by several of his students, he included much broader group factors such as arithmetic, mechanical, and linguistic abilities.

Finally, on the basis of a series of studies, additional *general* factors were suggested. These include *p* (perseveration), *o* (oscillation), and *w* (will), the last extending the theory to the field of personality traits. It was also proposed by Spearman (cf. 89, 91) that whereas *g* represents the total amount of "mental energy" at the subject's disposal, *p* may denote the inertia of such mental energy, and *o* the unsteadiness of its supply. Thus all the proposed general intellectual factors could be but different manifestations of the same *g* factor.

In the later writings of Spearman and his followers, the presence of all three classes of factors—general, group, and specific—is clearly recognized. The chief differentiating feature of the later form of the Two-Factor theory thus seems to be its relative emphasis upon the *g* factor as a more important influence than the group factors in producing correlation. It should also be noted that the distinction between general, group, and specific factors is probably not so fundamental as may at first appear. For example, if the number or variety of tests in a battery is small, a single "general" factor may account for all the correlations among them. But when the same tests are included in a larger battery with a more heterogeneous collection of tests, the original general factor may now appear to be only a group factor, common to some but not all the tests. Similarly, a certain factor may have occurred in only one of the tests in the original small battery, but may be shared by several tests in the larger battery. Such a factor would have been identified as a specific in the original battery, but would become a group factor in the more comprehensive battery. It is probably more realistic to speak of group factors of varying extent, rather than of sharply differentiated general, group, and specific factors.

The Sampling Theory. The Sampling theory of trait organization has been most clearly and completely formulated by Thomson (99) and is described by him in a series of publications dating from the second decade of the present century. According to such a theory, behavior depends upon a very large number of independent elements,

which have been variously identified by Thomson and others (102, 115) with genes, neural elements, stimulus-response bonds, specific experiences, or environmental characteristics. Any one activity of the individual, it is argued, depends upon a particular *sample* or combination of these elements. Correlation results from the overlapping of different samples of elements. Different types of factors may thus be produced, varying from the specific, through group factors of differing extent, to a very broad or general factor. Thomson has repeatedly illustrated, with data from dice throws,[2] how various factor patterns may occur from overlapping samples of independent elements.

Improvement in an activity with practice, according to Thomson's Sampling theory, is not due to improvement in the elementary abilities involved, but to the use of a more economical and efficient selection of these abilities. As a practical illustration of this, Thomson cites the well-known dropping out of unnecessary movements in the learning of a motor skill.

Other viewpoints which bear a fundamental resemblance to that of Thomson are those expressed by E. L. Thorndike (100, 106) and Tryon (115). Thorndike's views on trait relationships seem to have run the gamut from extreme specificity to the opposite extreme of a single general factor (cf. 100–106). Throughout his various statements, however, one can discern the conception of abilities as being ultimately reducible to a large number of simple associative bonds or connections, whose role in trait relationships appears to be quite similar to that of Thomson's elements. Tryon has expressed a similar view, in terms of the operation of a multitude of elementary psychological components. The overlap among such elementary components produces the correlations between different functions. As further, although minor, sources of such correlations he mentions possible associations between environmental fields and between gene-blocks. By the former he refers, for example, to the fact that the individual in an inferior cultural milieu may lack many environmental opportunities for developing *both* linguistic and computational skills. Cultural linkages would thus tend to produce a correlation between these two areas. Correlations between independent gene-blocks could occur through assortative mating. Since individuals tend to marry within their own general socio-economic and educational level, persons

[2] Frequently employed in statistics as a means of obtaining purely random or "chance" data.

superior in quite different respects are likely to interbreed. Their offspring would thus tend to receive genes for superior development in a number of initially unrelated characteristics. The same type of selection would operate in the interbreeding of persons of diverse inferiority.

The Multiple-Factor Theory. The theory held by the largest number of contemporary psychologists proposes a relatively small number of moderately broad group factors, each of which may enter with different weights into different tests. For example, the verbal factor may enter with a high weight into a vocabulary test, with a smaller weight into an analogies test, and with a very small weight into an arithmetic reasoning test. Such theories have been variously designated Multiple-Factor or Weighted Group-Factor theories.[3]

The publication in 1928 of Kelley's *Crossroads in the Mind of Man* (63) paved the way for a large number of studies in quest of particular group factors. Kelley contended, after a critical analysis of the methodology and data of Spearman, that the general factor is of relatively minor importance and can usually be attributed to the heterogeneity [4] of the subjects and to the common verbal nature of the tests employed. If a residual general factor should be found when these influences are ruled out, Kelley maintained that it would probably be small and insignificant. The major relationships among tests he attributed to a relatively small number of broad group factors. Chief among these were manipulation of spatial relationships, facility with numbers, facility with verbal material, memory, and mental speed. This list has been modified and extended by subsequent investigators, employing the more recent methods of factor analysis to be considered in the following section.

One of the leading exponents of the Multiple-Factor theory today is Thurstone (108–114). On the basis of extensive research by himself and his students, Thurstone has proposed about a dozen group factors which he designates "primary abilities." Those most frequently corroborated in the work of Thurstone and of other independent investigators include the following: [5]

[3] For a relatively early but clear exposition of the operation of weighted group factors, cf. 62, pp. 195–226.

[4] The influence of heterogeneity upon correlation coefficients will be discussed in the following section.

[5] For some of the most relevant investigations dealing with these factors, as well as for general summaries, cf. 3, 4, 31, 36, 42, 51, 52, 53, 83, 108, 111, 114, 126.

V: *verbal comprehension*—the principal factor in such tests as reading, verbal analogies, disarranged sentences, verbal reasoning, and proverb matching. It is most adequately measured by vocabulary tests

W: *word fluency*—found in such tests as anagrams, rhyming, or naming words in a given category (e.g., boys' names or words with the same initial letter). This factor has been identified in relatively few investigations.

N: *number*—most closely identified with speed and accuracy of simple arithmetic computations.

S: *space*—it is possible that this factor may represent two distinct factors, one covering the perception of fixed spatial or geometric relations, and the other "manipulatory visualization" in which changed positions or transformations must be visualized (51).

M: *associative memory*—found principally in tests demanding rote memory for paired associates. The evidence is against the presence of a broader factor through all memory tests (3, 4, 51, 108). Other restricted memory factors through narrowly defined groups of tests have been suggested by some investigations (51).

P: *perceptual speed*—quick and accurate grasping of visual details, similarities, and differences. This factor may be identical with the "speed factor" identified by earlier investigators (cf. 36, 126) and described as "speed in dealing with very easy material." It may also be restricted to visually presented material (51).

I (or *R*): *induction* (or *general reasoning*)—the identification of this factor is probably least clear. Thurstone originally proposed an inductive and a deductive factor (108). The latter was best measured by tests of syllogistic reasoning and the former by tests requiring the subject to find a rule, as in a number series completion test. Evidence for the deductive factor, however, was much more tentative than for the inductive. Moreover, other investigators suggest a general reasoning factor, illustrated by such tests as arithmetic reasoning, and fail to corroborate the distinction between inductive and deductive reasoning factors (51).

Rapprochement. The problem of trait organization has been one of the most controversial in psychology. In the 1920's and 1930's the journals fairly bristled with critiques, replies, rejoinders, and counter-rejoinders. The exponents of each point of view were clearly aligned on one side or another. With the gradual sharpening and clarification of concepts and the steady accumulation of relevant data, a rapprochement has been slowly but unmistakably occurring among

these points of view. It has already been noted above that the distinction between the Two-Factor and the Multiple-Factor theories today is but one of emphasis and degree. The three types of factors—general, group, and specific—are not sharply differentiated, but probably represent a continuum of factors of varying breadth.

A less obvious but equally fundamental convergence has occurred between the Multiple-Factor and the Sampling theories. On the one hand, multiple-factor exponents have agreed that their factors, rather than being unitary ultimates, may well represent aggregates of more elementary units akin to the elements of the Sampling theory. The discovery of a "verbal factor" might thus mean simply that a particular combination of response elements, all dealing with verbal material, was discernible. The term "*functional unity*," recently proposed by Thurstone (109) for such an aggregate, typifies this point of view. Its connotations appear to be essentially the same as those of Tryon's concept of "*operational unities*" among the elementary psychological components (116). Thomson, too, beginning from the opposite extreme, has drawn closer to the multiple-factor view (99). Although he originally assumed the sampling of elements by different functions to be completely random, he subsequently proposed that the elements are organized into fairly enduring "*sub-pools of the mind*." These sub-pools into which the elements are structured or organized would account for the correlations within each area, such as the verbal, numerical, or spatial. With the "primary abilities" of the Multiple-Factor theory broken down into functional or operational unities consisting of numerous uncorrelated elements, and with the elements of the Sampling theory organized into sub-pools, the rapprochement appears to be virtually complete.

In keeping with these revised concepts, too, is the point of view which Burt has been championing for many years (8, 19, 20). Factors, according to Burt, should be regarded as principles of classification or *descriptive categories* rather than as causal entities.[6] The factors cannot explain the test scores. Rather can it be said that the test scores account for the factors, since the latter are derived from the test scores. Essentially, factors represent a summary and simplified statement of the information contained in test scores, and therein lies their practical value. The function of statistically derived factors as a *simplification* of behavior descriptions will be more clearly appar-

[6] For another expression of the same point of view, cf. 7.

ent when we consider techniques of factor analysis in the following section.

METHODOLOGY

The Tetrad Criterion. Fundamentally, all techniques for the study of trait organization are based upon the correlation coefficient. This measure indicates the degree of relationship between two sets of scores, or the extent to which each individual's performance on one test corresponds to his performance on another test. Correlation, however, cannot analyze the *mutual interrelationships* of a large number of variables. A correlation coefficient may indicate whether there is some factor common to a pair of tests, but it cannot show the presence of a single common factor through three, or four, or any larger number of tests. Let us suppose that all intercorrelations among three tests have been computed, with the following results: [7]

$$r_{12} = .60$$
$$r_{13} = .49$$
$$r_{23} = .70$$

Although all three correlations are positive and rather high, we cannot determine whether these three tests have one common factor or several common factors among them. Test 1 might share one factor (A) with test 2, and a different factor (B) with test 3; a still different factor (C) might constitute the common element between tests 2 and 3.

It was Spearman (88) who first demonstrated that from the *relationships among correlation coefficients* it is possible to discover the factorial organization of any number of tests. The first method proposed by Spearman was the *hierarchical arrangement* of correlation coefficients. According to this criterion, if it was possible to arrange all the intercorrelations among a set of tests in such a way that they decreased consistently in size both along the rows and along the columns of the table, then the relationships among these tests could be explained entirely in terms of *g* and *s*. This was a relatively crude "inspectional" method of determining hierarchy. Subsequently, the computation of the *intercolumnar correlation* was suggested as a convenient numerical index of hierarchy. The intercolumnar correlation

[7] It is customary to denote the particular variables correlated by subscripts. Thus, r_{12} is the correlation between test No. 1 and test No. 2.

is simply the correlation between columns of correlation coefficients. A + 1.00 intercolumnar correlation would indicate a perfect hierarchical arrangement of the coefficients.

Finally, the intercolumnar correlation was replaced by the *tetrad criterion*. The latter gets its name from the fact that the tests are considered in sets of *four*. For every four tests, or variables, we can compute three tetrad equations as follows:

$$t_{1234} = r_{12} \times r_{34} - r_{13} \times r_{24}$$
$$t_{1243} = r_{12} \times r_{34} - r_{14} \times r_{23}$$
$$t_{1342} = r_{13} \times r_{24} - r_{14} \times r_{23}$$

Spearman and others have been able to prove mathematically that if all three tetrad equations are equal to *zero*,[8] then a single common factor is sufficient to account for the relationships among the four variables.

This was a decided step forward from the simple correlation coefficient. It was now possible to analyze the interrelationships of any number of variables by computing different sets of tetrads. The extension of the tetrad criterion beyond four variables can easily be demonstrated. Let us suppose that we have administered six tests to the same subjects. First, we compute the three tetrad equations with tests 1, 2, 3, and 4. If all three tetrads are equal to zero, we may conclude that the same factor which underlies tests 1 and 2 is also common to tests 3 and 4. Then if the tetrad criterion is likewise satisfied (i.e., all tetrads equal to zero) with tests 1, 2. 5. and 6, we know that the factor common to 1 and 2 is identical with that common to 5 and 6. Hence the same factor must be common to all six variables.

The development of the tetrad criterion stimulated extensive research on trait organization and was undoubtedly an important step forward in the statistical study of trait relationships. Its usefulness, however, is limited. One practical drawback is that, as the number of tests increases, the number of tetrads to be computed becomes excessively large. With only 10 tests, for example, there are 630 tetrads. With 60 tests—the number employed in one of Thurstone's studies—over a million tetrads would have to be computed. The technique thus becomes unwieldy in the analysis of a large number of tests. Moreover, the tetrad technique does not provide a clear

[8] Or are sufficiently close to zero to be within the error of sampling.

over-all picture of the location of group factors, nor does it indicate the weight with which these factors enter into each test. With the shift in emphasis from a single general factor to a number of group factors, the tetrad criterion has been largely supplanted by more suitable and expeditious methods of factor analysis.

Factor Analysis. Several alternative procedures for analyzing a set of test scores into their constituent factors have been developed by Kelley (64), Hotelling (61), Burt (19), Holzinger (59), Tryon (116), Thurstone (112), and others.[9] Although differing in their initial postulates, most of these methods yield results which are not too unlike. The most widely used technique is the *centroid method* developed by Thurstone (112). In common with all the other methods, this technique begins with a complete table of all the intercorrelations among a set of tests. Such a table is known as a *correlation matrix*. In the process of factor analysis, this correlation matrix is transformed into a *factor matrix*, giving the weight or loading of each factor in each test.

An example of such a factor matrix will be found in Table 28. This matrix was derived from the intercorrelations of 21 tests given to 437 seventh and eighth grade school children (114). The seven factors listed at the top of the table are the same as those described on page 497 and are indicated by the same letters. The entries in the body of the table show the loading of each test with each of these factors. For example, Test 1, *identical numbers*, has significant loadings of .42 and .40 with the perceptual speed and number factors, respectively; Test 8, *vocabulary*, has a high loading of .66 with the verbal comprehension factor and negligible loadings with all the other factors; Test 16, *addition*, has a single significant loading of .64 with the number factor. All the residuals, given in the last column of the table, are small, indicating that substantially all the correlation among the tests can be accounted for by the seven factors shown.

Rotation of Axes. The factor weights found by the centroid method represent a "center of gravity" or a sort of average value based on all the correlations. As each factor is extracted, the residual correlations are subjected to the same type of analysis in order to obtain the weights of the next centroid factor. Thurstone (109, 112)

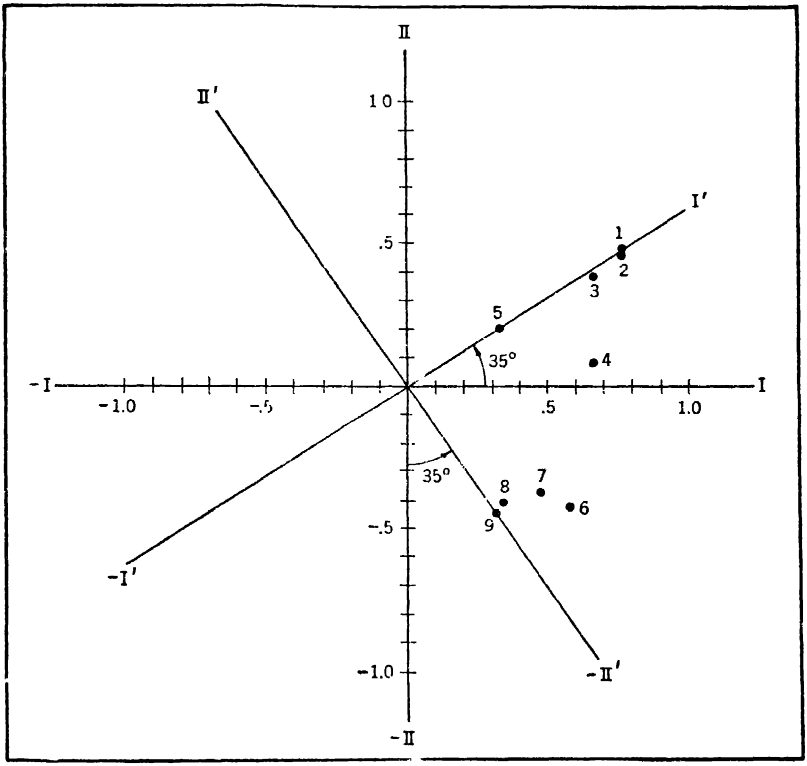
[9] For a survey of the assumptions underlying the various methods and a clear introduction to techniques of factor analysis, cf. Wolfle (126) and Guilford (48, Ch. XIV).

TABLE 28 *A Factor Matrix Based upon the Scores of 437 School Children on 21 Tests*

(From Thurstone and Thurstone, 114, p. 91)

Tests	*Factor Loadings*							
	P	*N*	*W*	*V*	*S*	*M*	*R*	Residual
1. Identical Numbers	.42	.40	.05	−.02	−.07	−.06	−.06	.08
2. Faces	.45	.17	−.06	.04	.20	.05	.02	−.12
3. Mirror Reading	.36	.09	.19	−.02	.05	−.01	.09	.12
4. First Names	−.02	.09	.02	.00	−.05	.53	.10	.02
5. Figure Recognition	20	−.10	.02	−.02	.10	.31	.07	−.17
6. Word-Number	.02	.13	−.03	.00	.01	.58	−.04	.04
7. Sentences	.00	.01	−.03	.66	−.08	−.05	.13	.07
8. Vocabulary	−.01	.02	.05	.66	−.04	.02	.02	.05
9. Completion	−.01	.00	−.01	.67	.15	.00	−.01	−.11
10. First Letters	.12	−.03	.63	.03	−.02	.00	−.00	−.08
11. Four-Letter Words	−.02	−.05	.61	−.01	.08	−.01	.04	−.05
12. Suffixes	.04	.03	.45	.18	−.03	.03	−.08	.10
13. Flags	−.04	.05	.03	−.01	.68	.00	.01	−.07
14. Figures	.02	−.06	.01	−.02	.76	−.02	−.02	.07
15. Cards	.07	−.03	−.03	.03	.72	.02	−.03	.13
16. Addition	.01	.64	−.02	.01	.05	.01	−.02	−.03
17. Multiplication	.01	.67	.01	−.03	−.05	.02	.02	.01
18. Three-Higher	−.05	.38	−.01	.06	.20	−.05	.16	−.12
19. Letter Series	−.03	.03	.03	.02	.00	.02	.53	.02
20. Pedigrees	.02	−.05	−.03	.22	−.03	.05	.44	−.02
21. Letter Grouping	.06	.06	.13	−.04	.01	−.06	.42	.06

has argued that the factors thus located [10] do not usually correspond to meaningful categories, and he therefore advocates the rotation of axes following the centroid factor analysis. The factorial matrix reproduced in Table 28 has already been rotated.



Original Centroid Axes I, II Rotated Orthogonal Axes. I', II'

Fig. 83. Rotation of Axes: Orthogonal. (From Garrett, 42, p. 261.)

Factors can be visualized geometrically as axes in terms of which each test can be plotted or described. Reference to Figure 83 will make this interpretation of factors clearer. For simplicity of illustration, only two factors, or axes, are included in this figure. Three factors would require a three-dimensional representation, and four or more factors could not, of course, be represented directly in geo-

[10] Either by the centroid method or by other similar techniques of factor analysis.

metrical space, although they can be handled conceptually and mathematically by an extension of the same principles. In Figure 83, the factor loadings of the following nine tests have been plotted with reference to the two centroid axes, I and II:

1. Vocabulary
2. Opposites
3. Analogies
4. Sentence Completion
5. Disarranged Sentences
6. Arithmetic Reasoning
7. Number Series Completion
8. Equation Relations
9. Multiplication

The factor loadings were found from the intercorrelations of the scores of 210 college men on these nine tests (83, 42). Reference to Figure 83 will show, for example, that Test 1, *vocabulary,* has a loading of .776 with Factor I (distance along horizontal axis) and a loading of .479 with Factor II (distance along vertical axis). Thus by plotting each test we have "described" and located it with reference to the two centroid axes.

It will be noted that nearly every test in Figure 83 has an appreciable loading with both factors and that several negative weights are present. A number of psychologists regard negative loadings as "psychologically meaningless," [11] because such loadings imply that the higher an individual rates in the particular factor, the poorer will be his performance on the test. Centroid factors are also difficult to describe in psychological terms. The two centroid axes in Figure 83, for example, do not lie close to any cluster of tests through which the factors might be identified. It was as a solution for these difficulties that Thurstone proposed the rotation of axes after factorization (109).

It should be noted that a given set of tests can be described in terms of an infinite number of coordinate systems or axes. The situation is not unlike that of locating points in geography (125). Longitude *could* be measured from Chicago instead of Greenwich as a point of origin, and latitude *could* be measured from the North Pole instead of the Equator. In factor analysis, the choice of axes depends largely upon the objectives of the analysis. Thurstone has proposed what he terms "*simple structure*" as a criterion for locating the most suitable axes for the analysis of psychological tests. By this he means:

[11] At least in the measurement of abilities; the possibility of genuine negative loadings seems more acceptable in connection with personality traits.

(a) eliminating negative factor loadings as completely as possible, and (b) maximizing the number of zero (or near-zero) factor loadings. The latter condition implies that each test shall be described by the smallest possible number of factors, since a zero loading simply means that the factor does not enter into that particular test. Thurstone maintains that when the axes are rotated in such a way as to maximize the zero loadings, most negative loadings also disappear and the resulting factors are relatively easy to interpret or identify in psychological terms. Referring back to Figure 83, we may note that the rotated axes I′ and II′, which fulfill the conditions of simple structure fairly closely, seem to correspond to the familiar *V* and *N* factors. Rotated axis I′ runs close to a cluster of verbal tests (Tests

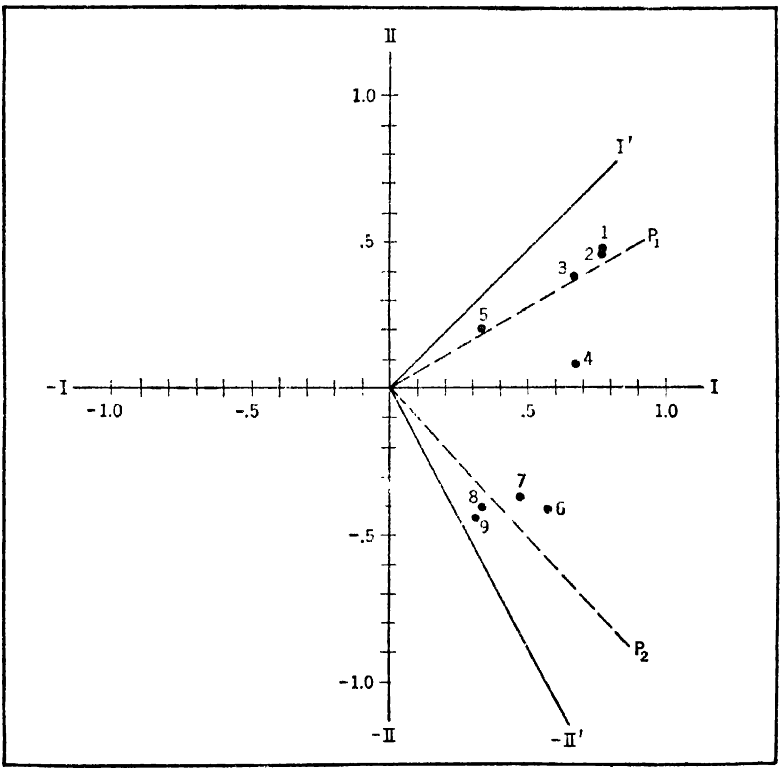


Fig. 84. Rotation of Axes: Oblique. (From Garrett, 42, p. 262.)

1 to 5), while axis II′ runs close to the numerical tests (Tests 6 to 9).

It will be observed that in Figure 83 *both* axes were rotated over an angle of 35°. The two axes thus remained at right angles to each other, or *orthogonal*. An even closer approximation to simple structure can be reached if each axis is allowed to rotate independently of the other, as shown in Figure 84. In this case, the new axis I′ is set at right angles to the line P2, which runs through the cluster of numerical tests. Thus the numerical tests have close to zero loadings in Factor I′. Similarly, the new axis II′ is set at right angles to the line P1 through the verbal cluster. The two new reference axes, I′ and II′, are not orthogonal to each other, but represent *oblique axes*. This signifies that the two factors which have been identified are themselves correlated. In the example illustrated in Figure 84, the *V* and *N* factors were correlated to the extent of .225.

There is an increasing recognition of the fact that oblique axes, or correlated factors, may be just as useful as orthogonal axes in a systematic description of behavior. Logically, there is no reason why the primary categories of behavior *must be* uncorrelated. In physical measurement, for example, height and weight have clearly demonstrated their usefulness as bodily dimensions, despite the fact that they are certainly correlated (109).

One last point should be noted regarding the rotation of axes. If the factors themselves are correlated, then it should be possible to "factorize the factors" and locate *second-order factors*. This has been done in an analysis of the scores of 710 eighth grade children on 60 tests (114). A single second-order general factor was identified which seemed to be fairly similar to Spearman's *g*. Such a finding should perhaps be viewed as further evidence of rapprochement among the various theories described in the preceding section.

Some Limitations of Factor Analysis. In interpreting the results of factor analysis, certain major limitations of these techniques should be considered. First, any set of intercorrelations can be analyzed into innumerable factor patterns. To demonstrate the presence of certain factors simply shows that the tests *can* be described in terms of those factors, not that they *must*. An analysis into other factor patterns is not precluded. To illustrate this point, Spearman (90) cited the analogy with a rectangle. Such a figure can be divided into two triangles, but it can also be divided in an infinite number of other ways.

How we divide a rectangle or how we factorize a test battery depends upon the nature of the problem. The various methods of factor analysis differ in the limiting conditions which they impose in order to reach a determinate solution. The appropriateness of these conditions to the problem under investigation should determine the choice of method

Secondly, since all techniques of factor analysis begin with intercorrelations, it is obvious that any circumstances which affect the correlation coefficient will also affect the results of the factor analysis. It has been repeatedly demonstrated, both empirically and theoretically,[1-] that the size of a correlation coefficient is affected by the *heterogeneity* of the group of subjects upon whom the data were collected The most obvious example is that of age heterogeneity. If the subjects range in age from 3 to 15 years, a high positive correlation will be found between even such diverse characteristics as size of the great toe and Stanford-Binet mental age The same two measures would yield a zero correlation within a homogeneous age group such as, for example, 10-year-old children.

Nor does heterogeneity always raise the correlation coefficient; it may lower it Let us suppose, for example, that a group of high school boys and a group of high school girls have each taken two memory tests, one based on a sports story, the other on a fashion story Let us further suppose that the correlation between the two memory tests is 40 among the boys and also .40 among the girls The girls as a group, however, will probably score higher than the boys on the fashion test while the boys will excel on the sports test. If, now, we compute a single correlation between these two tests in the combined group of boys and girls, the resulting coefficient will be much lower than 40 The greater the sex difference on the two tests, the more will the correlation be lowered by combining the two groups

It is likewise possible for heterogeneity to produce a negative correlation between two variables which are actually uncorrelated Thus, if a heterogeneous group composed of Chinese and Scandinavians were rated for height and for proficiency in the use of chopsticks, a fairly high negative correlation would be obtained between these two measures. The Chinese would, in general, be shorter than the Scandinavians and definitely more adroit with chopsticks Within

[1-] Cf e g 32 44

either group, however, we should scarcely expect any correlation between these two characteristics.

Correlations which result from a marked degree of heterogeneity in the group are usually regarded as *spurious correlations.* It is difficult to decide, however, just what constitutes a permissible degree of heterogeneity. Obviously, all heterogeneity should not be eliminated, even if this were possible, since individual differences would thereby disappear and correlation would be meaningless. The desired degree of heterogeneity must be determined on the basis of the particular problem under investigation. It should always be remembered, however, that correlation coefficients, or any statistical measures derived from them, must be interpreted in the light of the particular group upon which they were obtained.

APPLICATIONS OF FACTOR ANALYSIS

Special Areas of Ability. Factor analysis began as a technique for studying the organization of intelligence. As a method, however, it is proving to be applicable to an increasing number of widely diverse questions. Intensive studies of special areas of ability have been under way to supplement the broad surveys of the earlier investigations. Studies of *learning* among adult subjects, for example, have so far failed to reveal a general learning factor (127, 128). Gains tend to be specific: the same individual may be a relatively rapid or good learner in one task and a slow or poor learner in another.

A factor analysis of *perception,* with both paper-and-pencil and laboratory tests, disclosed several significant factors within this area (110). Among them were reaction time, speed of perception, speed of judgment, rate of reversals in ambiguous figures, speed of closure, and flexibility of closure.[13] The last two are of special interest because of their suggested association with certain intellectual and personality characteristics. For example, a survey of a group of administrators in Washington showed that the more successful administrators scored relatively well on some of the closure tests (110). A tentative hypothesis, suggested to account for these results, was that the successful

[13] "Closure" is a term introduced in the Gestalt studies on perception to refer to the well-known perceptual filling of incomplete figures. Thus a sketch of a house will be perceived as a complete house even though the drawing itself may contain many gaps.

administrator is an individual who can most readily unify the apparently unrelated elements in the work which he must coordinate. That the previously identified perceptual speed factor (*P*) can itself be broken down into a number of subsidiary factors has been demonstrated in more than one study (13, 110).

In a special study of *verbal* tests, the previously identified factors of verbal comprehension (*V*) and word fluency (*W*) were further subdivided into three and two factors, respectively, and three new verbal factors were isolated (24). Thus eight new factors of more restricted scope than the original *V* and *W* were identified in verbal tasks. Examples of these factors include the individual's stock of linguistic responses, facility and fluency in oral speech, and speed of articulatory movements. A factorial analysis of *fluency in writing* (95) also suggested new factors, such as "ideational fluency" and "verbal versatility." Intensive factorial investigations of *reasoning tests* have likewise suggested the presence of several uncorrelated reasoning factors in such tests (34; 119, No. 5). Whether any one of these factors is common to *all* reasoning tests is still a moot point. Attempts to identify a factor of *flexibility*, or the ease with which the individual can shift from one task to another, have so far failed to disclose any such factor; the subject's ability to shift seems to depend entirely upon the specific content of the tests (65).

Non-Intellectual Functions. Some application of factor analysis has also been made to non-intellectual functions. A number of investigations have been concerned with *motor functions*, including both manipulatory skills and athletic proficiency (16, 21, 38, 66, 86, 87). Such functions have on the whole proved to be highly specific, the intercorrelations among different motor tests usually being quite low. Certain relatively narrow factors have, however, been identified. Steadiness tests, for example, have repeatedly shown a common factor. There is also some evidence for group factors underlying improvement in motor functions, speed of isolated reactions, finger and hand speed in restricted oscillatory movements, and forearm and hand speed in restricted oscillatory movements. In general, the factors most commonly found in motor functions seem to be related not to the specific muscle groups, body parts, or sensory modalities involved, but rather to similarities in pattern or type of movement (86). Finally, it should be noted that the more complex motor tests often contain the previously identified space factor (*S*). Moreover, several

analyses of common mechanical aptitude batteries show them to involve principally the space factor (*S*), perceptual speed (*P*), and various motor factors (47, 58, 125).

Factorial studies of *sensory tests* have also been undertaken, including visual acuity and other widely used measures of visual efficiency (118, 124). One finding of practical significance in this connection is that different visual tests designed to measure the same characteristics may have different factorial compositions. As a result, such tests will not be equally valid for different purposes. Among non-intellectual functions, mention should also be made of the rapidly growing application of factorial methods to *personality measurement*. Some of the problems and results in this area will be considered in a subsequent section, but it might be noted that factor studies have been contributing to the breakdown of the traditional demarcation between personality and intelligence. The same test may measure factors in both categories. Nor are personality and ability variables wholly unrelated within the individual. There is some evidence, for example, which suggests a relationship between ability in drawing and certain personality characteristics (27).

Academic and Vocational Areas. A further application of factorial techniques is represented by the analysis of performance in various academic and vocational areas. This use of factor analysis has important practical implications for personnel selection and counseling, since it places the construction of aptitude test batteries on a firmer and more systematic foundation. Let us suppose that we want to discover what factors are involved in successful performance in courses in elementary French or calculus, or in the occupations of filing clerk, cabinet maker, or city editor. The procedure would be to assemble a trial battery of tests sampling all the major factors and then to factorize the criterion measure along with the test battery. Thus final grades and achievement test scores in French or calculus, or follow-up records of job performance, would be included as one variable in the correlation matrix. The factor matrix will then show the loading of the criterion with each factor. For example, if variable 1 in Table 28 had been a criterion measure rather than a test score, we could find the contribution of each of the seven factors to this criterion by simply reading across the first row of the table. The next step would be to choose those tests in the battery which are most heavily weighted with the factors that predominate in the criterion.

These are the tests that will be most effective in predicting successful performance in the educational course or occupation under consideration.

This type of analysis has been conducted with performance in such scholastic areas as algebra (15, 78), geometry (60), and technical courses (35). A good example of its use in military psychology is provided by the factorial analysis of pilot performance conducted by the psychological staff of the AAF (37, 51, 52, 53, 119). In this project, nearly 30 factors were identified, covering abilities, interests, emotional characteristics, and educational and other background variables. Factor analysis likewise constituted the basic technique followed by the United States Employment Service in devising its General Aptitude Test Battery (40, 92). Preliminary batteries consisting of 15 to 29 tests were administered to nine groups totaling 2156 men between 17 and 39 years of age. Most of the men were trainees in vocational courses. The 10 factors identified most clearly and incorporated into the U.S.E.S. battery have already been cited in the preceding chapter. They include:

G — general intelligence	*Q* — clerical perception
V — verbal ability	*A* — aiming
N — numerical ability	*T* — motor speed
S — spatial ability	*F* — finger dexterity
P — form perception	*M* — manual dexterity

Miscellaneous Applications. The possible uses of factorial techniques in psychology and related fields are many and varied. The factorial analysis of bodily dimensions in the study of constitutional types has already been mentioned in Chapter 13. Other proposed applications range from the classification of psychiatric syndromes and the simplification of scratch tests for allergy to the analysis of voting records, Supreme Court decisions, and stock market fluctuations (109, 113). In most of these areas, exploratory research has already begun. Factor analysis is also being currently employed as a technique for simplifying "job evaluation" systems in business and industry (67).

In certain applications of factor analysis, an adaptation known as obverse or *inverted factor technique* has been employed.[14] This simply

[14] Burt has pointed out that, strictly speaking, this technique is based upon a *transpose* rather than upon the *inverse* of the usual matrix of measurements, since the rows are written as columns (19, p. 169). For a discussion of the technique, cf. 19, Ch. VI; 93.

means that the original correlations are correlations between persons rather than between tests. Thus all the scores of individual A on, let us say, 30 tests are correlated with the scores of individual B on the same 30 tests, yielding r_{AB}. Similar correlations are found for every other pair of individuals in the group. These intercorrelations then form the basis for a factor analysis by any of the usual techniques. Inverted factor analysis has been proposed especially as a means of investigating personality types, since the "group factors for persons" would then represent "type factors" or patterns of traits shared by certain individuals. In some situations, as when an extensive series of measures is available on a relatively small number of persons, inverted factor technique may be preferable. It does not seem, however, that the two approaches should be regarded as fundamentally different. Substantially the same factors would probably be found by either approach (19).

GROUP DIFFERENCES IN FACTOR PATTERNS

With the extension of factor studies to subjects differing in age, sex, education, occupational background, and other characteristics, certain consistent group differences have come to light.[15] What at first appeared as a source of confusion and controversy is now gradually falling into a systematic picture. Through the comparison of factor patterns in diverse groups, moreover, we may learn something about the nature of trait relationships and how traits develop. As early as 1927, Spearman called attention to such group differences, stating, "Another important influence upon the saturation of an ability with *g* appears to be the class of person at issue" (89, p. 217). At that time he also reported some data suggesting that among older as well as among brighter individuals, abilities are more specialized and the general factor plays a relatively smaller part. It is interesting to note that a large number of the studies by Spearman and his students were conducted on school children, a fact which may partly account for the insistence of these investigators upon the importance of the *g* factor. Most of the early studies by the American group-factorists, on the other hand, were concerned with college students. The latter found little or no evidence of a general factor, and put the major emphasis upon a few broad group factors.

[15] For a more detailed survey of these differences, cf. 9.

Age. A number of independent investigations are now available which indicate that abilities do in fact become more specialized as the child grows older (43). Among preschool children, the general factor appears to be relatively large, and group factors less important. For example, in a study on 200 5- and 6-year-old children (17, 42), various memory tests were as closely related to a vocabulary test and to Stanford-Binet MA as they were to each other. In contrast to this, at the college level simple tests of associative memory present a distinct group factor which breaks off sharply from *V*, *N*, and other group factors (3, 4, 108). Thus the correlation between vocabulary and the entire memory battery was .45 among the preschool children, but only .06 among the college students (43).

Similarly, in a re-analysis of two different studies (82, 83), Garrett (42) found a correlation of .83 between the *V* and *N* factors in a group of third and fourth grade school children, in contrast to a correlation of only .23 among college students. In the Thurstones' extensive study of 710 eighth grade school children with 60 tests (114), much higher correlations were found among the group factors than had been found in the earlier study on college students by one of the authors (108). For example, the *N* factor correlated .33 with word fluency; *I* correlated .43 with *S* and .42 with *V*. The correlation of *V* with *W* was .42 and with *S* .38. In the college sampling, all factorial correlations were negligible, the median correlation being .03 and the highest .24 (108, p. 100). In the eighth grade sampling, furthermore, a second-order general factor was identified whose correlations with the first-order group factors ranged from .14 (with *M*) to .72 (with *V*).

A few studies have been specifically designed to discover the role of age in trait relationships. In one of these (45), three groups of school children, aged 9, 12, and 15, respectively, were given tests of memory, verbal, numerical, and spatial aptitudes, and motor speed. The intercorrelations among these tests tended to decrease from the youngest to the oldest group. Factor pattern analyses revealed a general factor whose average contribution to the total battery dropped with age. Among the boys, the average per cent contribution of the general factor was 31, 32, and 12 for ages 9, 12, and 15, respectively. For the girls, the corresponding per cents were 31, 24, and 19. These results were corroborated in a later study using the Thurstone Tests of Primary Mental Abilities with 11-, 13-, and 15-year-old

boys. The latter study confirmed Thurstone's findings of a second-order general factor, and indicated that the influence of this factor drops with age (29). In still another study (11), a single group of children was retested, the average age at the two testings being 9 and 12. Eight tests covering verbal, numerical, and spatial content were administered. Intercorrelations dropped from the first to the second testing, the decrease being larger in the correlations *between* verbal and numerical tests than those within either group. Factor pattern analyses corroborated the findings of other studies: a large general factor was found at both age levels, but its magnitude dropped from age 9 to age 12.

The standardization data of the Wechsler-Bellevue Intelligence Scale provide some information regarding age changes in factor pattern among adults (12). The average intercorrelation of the subtests in this scale dropped steadily from the 9-year-old group to the 25–29-year-old group, thus corroborating the results of other studies. In the 35–44-year group, however, it rose to .31, and in the 50–59-year group it rose again to .43. Factor analyses showed evidence of a predominant general factor in the 9-year group and again in the age group 50–59, while in the intervening ages group factors played the major part. Thus in this study, specialization seemed to reach a peak during the middle age levels; in both the younger and the older groups, generalization of ability seemed to be the rule.

One of the first hypotheses to account for age changes in factor patterns was that proposed by Kelley (63). The presence of a general factor in childhood, according to this explanation, results from individual differences in rate of intellectual maturation. Thus the child whose mental development is slower would have relatively low scores on all the tests, while the faster developing child would have higher scores throughout. By an extension of the same hypothesis, the increasing weight of the general factor beyond maturity might be attributed to individual differences in the rate of mental decline. One objection to such a hypothesis comes from our present knowledge of intellectual growth. It will be recalled from Chapter 9 that different functions are quite independent in their development, and that it is unlikely that the individual is characterized by a general "rate of growth" or "rate of deterioration" for all abilities. Moreover, age changes need not *imply* maturational processes. We cannot assume that the same changes would occur regardless of what individuals did

during those years. That the latter is in fact important is suggested by some of the data to be considered in the following sections

Education. In discussing some of the later developments of his Sampling theory, Thomson wrote " . . a general tendency is noticeable in experimental reports to the effect that batteries do not permit of being explained by as small a number of factors in adults as in children, probably because in adults education and vocation have imposed a structure on the mind which is absent in the young . . . Some of this 'structure' is no doubt innate, but more of it is probably due to environment and education and life' (99, pp 306, 319) What is the evidence for the influence of education in this increasing 'structuring' of abilities from childhood to maturity?

First it should be noted that in all the studies on school children and college students discussed above *the older groups invariably had more education* Thus the 15 year-olds have had more education than the 9-year-olds, and the college students of course have had more than any other group Even more cogent is the fact that, in the Wechsler Bellevue data changes in factor patterns among older persons closely paralleled educational differences The 25 29-year group, showing the greatest specialization of ability, also had the highest education, with a range from one to four years of high school. The 35 44 year group, which ranged in education from the sixth grade to the first year of high school, showed less specialization of ability The oldest group with the least specialization, ranged in education from the fifth to the eighth grade Any of these changes or group differences in trait organization could thus be explained equally well in terms of education or age As long as both variables are present, we cannot choose between them on the basis of such data alone

What is needed is a comparison between different age groups of the same education or between different educational groups of the same age Some relevant preliminary data of this sort were provided by the army testing in World War II Within a sampling of 5000 men, carefully chosen so as to be representative of the entire army, intercorrelations were computed among the sub-test scores of the AGCT (Form 3a) and of the Army Mechanical Aptitude Test (77) The average age of this group was 27, and their average education 9½ grades In Table 29 will be found the intercorrelations of the parts

of the AGCT. It should be noted that Form 3a of this test is especially suited to such a correlational analysis since, unlike the shorter forms, it consists of separate sub-tests, each of which is timed separately.

TABLE 29 *Intercorrelations among the Sub-Tests of the AGCT in a Random Sample of 5000 Cases*

(Data from Personnel Research Section, A G O , 77)

Tests	*2*	*3*	*4*
1. Reading and Vocabulary	.81	.81	.71
2. Arithmetic Computation		.90	.73
3. Arithmetic Reasoning			.75
4. Pattern Analysis			

Table 30 shows the intercorrelations of the parts of the Army Mechanical Aptitude Test with each other and with total AGCT scores.

These correlations are much higher than those found for similar tests among college students. But even more conspicuous is the relative uniformity of the correlations, regardless of test content. Such uniformity suggests that the relationships could be expressed in terms of a single general factor. The tetrad criterion, for example, would be readily satisfied when all correlations are nearly alike. Especially interesting are the correlations of the three mechan-

TABLE 30 *Intercorrelations among Army Mechanical Aptitude Sub-Tests and AGCT Total Score in a Random Sample of 5000 Cases*

(Data from Personnel Research Section, A G O , 77)

Tests	*2*	*3*	*4*
1. Mechanical Information	67	.78	.77
2. Surface Development		71	.76
3. Mechanical Comprehension			77
4. AGCT-3: Total Score			

ical aptitude sub-tests. It will be seen in Table 30 that these tests correlate .67, .78, and .71 with each other, and they correlate .77, .76, and .77 with AGCT total scores. The correlations of the mechanical tests with the separate sub-tests of the AGCT were nearly as high, ranging from .65 to .72. From an examination of such cor-

relations alone, it would be impossible to pick the correlations between two mechanical tests and those between a mechanical and a verbal or numerical test, since they are all so nearly alike. To be sure, it is quite probable that all these correlations were spuriously raised by inadequate control of testing conditions. Thus if a given individual was incapacitated by illness, fatigue, or other physical discomfort, his scores on all parts of the AGCT would be lowered by about the same amount. If, moreover, the Mechanical Aptitude Test and the AGCT were given within a short time of each other, the same disturbing condition might affect performance on both tests. Any uncontrolled factors in test administration, such as distractions or improperly given directions, would likewise tend to raise or lower the scores of a particular group on all parts of a test, thus raising the intercorrelations and making them more uniform. It is doubtful, however, whether such spurious factors could account for the major part of the obtained correlations. It seems reasonable to expect that even if such factors had been controlled, the intercorrelations in Tables 29 and 30 would still be much higher than those found among similar tests given to college groups.

What such findings suggest is that adults whose educational level is no higher than that of children resemble children much more than they do college students in their trait relationships. Among persons of lower educational levels, irrespective of age, abilities appear to be less highly differentiated and the general factor is relatively conspicuous.

Sex. Some data on sex differences in trait relationships are also available. In an early English study (72) on mechanical aptitude in school children, for example, the boys' scores on the various *spatial* tests correlated more highly with each other and less highly with estimates of "general intelligence" than did the girls' scores. The author suggested that spatial tests depend more largely upon a special aptitude among boys, and are more largely influenced by "general intelligence" among girls. In a later study (74) on 7 year-old English school children, a factorial analysis identified a spatial factor among the boys but not among the girls. This investigator likewise concluded that some of the tests involving a space factor for the boys were performed by the girls "by means of their general intellectual facility."

Corroborative data are furnished by studies on American school children at ages 9 and 12 (11, 82). In these groups, the correlations

of the spatial tests with verbal and numerical tests were higher among girls than among boys. In the same groups, the intercorrelations of *verbal* tests with each other tended to be higher among girls, while the intercorrelations of *numerical* tests with each other were higher among boys. These sex differences in correlations were larger in the 12- than in the 9-year-old group, and thus seem to become more conspicuous with age. Several factor pattern analyses of *mathematical aptitude* and of performance in mathematics courses have yielded factors which differ in both number and nature for the two sexes (15, 78). Studies of *memory*, conducted with 9-, 12-, and 15-year-old children and with college students, suggest that memory operates more nearly as an independent trait among women than among men (3, 4, 45). The intercorrelations of memory tests with each other tended to be higher among women; at the same time, the community between memory and non-memory tests tended to be greater among men.

Such data suggest that those groups which excel in performance within a given area exhibit a more closely knit organization of performance within that area. In the studies cited, for example, the women excelled in average scores on the memory tests and also showed higher intercorrelations among such tests than did the men. In mechanical aptitude, on the other hand, the men excelled in level of performance and showed higher intercorrelations among such tests than did the women. Moreover, the correlations between spatial and non-spatial tests were lower among men than among women. It is possible that the same conditions which make for good performance along certain lines tend also to unify and crystallize such performance into a distinct "trait."

Other Group Differences. Factor patterns among different *occupational* groups also offer interesting fields for research. Very little information is available in this area, although a few investigations offer promising leads. For example, among groups of adult men, the intercorrelations of three manual dexterity tests were consistently higher among operatives in repetitive tasks than among clerks or skilled trades workers (9, 96). The average correlations were .41, .26, and .25, for operatives, clerks, and trades workers, respectively. On the other hand, the dexterity tests correlated lower with spatial tests among both operatives and trades workers than among the clerks. From such findings, one might speculate regarding the possible

role of motor and mechanical experience acquired during the vocational training and actual job performance of these three groups. A difficulty encountered in making comparisons among most vocational groups is that such groups generally differ in educational level, age, and other variables.

Differences in factor patterns among *cultural groups* ought also to be considered. Relevant data are virtually non-existent in this area. The reason is undoubtedly to be found in the difficulty of devising a battery of tests applicable to widely diverse cultural groups. Although attempts have been made to construct tests which are relatively "culture-free," such tests do not offer sufficient variety and breadth of content to permit factor pattern analyses of the scope conducted within our culture. An analysis of the intercorrelations of scores on even such limited tests, however, would be illuminating. It would be very surprising indeed if in cultures very unlike our own we should find the emergence with age of the verbal, numerical, spatial, and other familiar "aptitudes" of our factor studies. To be sure, a certain degree of differentiation into relatively unified behavior traits may occur with age in all cultures. But the nature of such traits and the degree of differentiation are probably most unlike those found in our own culture.

Finally we may examine briefly the results of factorial studies on *infrahuman* groups. About a dozen investigations have been reported, most of them on white rats, but their results have all but defied interpretation.[16] In one study, for example, the intercorrelations of performance in nine tests were so low that nothing could be concluded beyond extreme specificity (71). As for the rest, the interpretations of the factors identified are highly speculative. Most of these factors are limited to a particular type of situation or learning problem. Some have been defined in terms of specific techniques which the animal may use in solving more than one problem, such as the principle of turning alternately right and left (107, 120). One is impressed, moreover, with the frequency with which factors related to emotional aspects of behavior appear in conjunction with "intellectual" factors. The relatively greater prominence of such emotional factors is also noteworthy. Among the factor descriptions, for example, can be found such terms as "a combination of intelligence and tameness,"

[16] Cf. 14, 39, 46, 71, 79, 84, 107, 117, 120, 121, 122, 123.

"wildness or panicky behavior," "wildness-timidity," and "self-confidence."

The difficulty of identifying factors in these animal studies, as well as the closer intertwining of "ability" factors with "emotional" factors than in the human studies, is not surprising when we consider certain facts about the subjects' backgrounds. White rats have not been subjected to formal education with standard sequences of courses in Elementary Maze Running 1–2, Problem Solving 5–6, or Advanced Seminar in String Pulling! Unlike the school children or college students of the human factor studies, the animals have not been exposed to that classic dichotomy between curricular and extracurricular, between standardized intellectual development and unstandardized emotional development. It has also been suggested that the inclusion of individuals from genetically different strains within the same group may account for some of the confusion and inconclusiveness of these factor analyses of animal behavior (84). Probably the most fruitful contribution that animal studies can make to the analysis of trait relationships is the experimental investigation of *how* factor patterns may be developed and altered in animals living under controlled laboratory conditions. The opportunities provided by this approach have scarcely been recognized.

TRAITS OF PERSONALITY

Typical Findings of Factor Analysis. The application of factorial techniques to the measurement of personality, although relatively recent, has come to represent an active and prolific area of research. Two of the best factorial analyses of personality *questionnaires* are those of Mosier (75) and the Guilfords (54, 55, 56, 57). In such studies, the initial data are the intercorrelations among individual *items*, rather than among test scores.[17] The subsequent procedure is the same as in factorial analyses of abilities, with the exception that negative factor loadings are not usually excluded in the rotation of axes. Since many personality traits may be regarded as "bipolar" (e.g., ascendance-submission, introversion-extroversion), negative loadings are more intelligible in this area than in the factorization of abilities.

[17] Since the responses to such items are generally twofold (Yes or No), a type of correlation known as *tetrachoric* is employed for this purpose, in place of the more familiar Pearson Product-Moment Correlation.

Mosier (75) administered 39 of the most discriminative items of the Thurstone Neurotic Inventory to 500 college men. Rather than finding emotional instability to be a unitary characteristic, Mosier found evidence for eight orthogonal traits in the responses to these items. The list, with tentative trait designations and illustrative behavior, is as follows:

C. Cycloid tendency: ups and downs in mood.
D. Depression: lonely, frequently in low spirits.
H. Hypersensitivity: feelings easily hurt.
I. Inferiority: lack of self-confidence.
S. Social introversion: shy, keeps in background on social occasions.
P. Public self-consciousness: difficulty in public speaking, stage fright.
Co. Cognitive defect: personality difficulties caused by individual finding himself intellectually below average of group in which he is placed.
Au. Autistic tendency: daydreaming, shut-in tendencies.

The interpretation of the last two factors, *Co* and *Au*, was much less clear and is offered very tentatively.

In a series of investigations by a similar method, the Guilfords (54, 55, 56) analyzed the most frequently recurring items in several introversion-extroversion questionnaires. This analysis was later extended to other types of personality questionnaires (57, 69). In the entire series, a total of thirteen factors [18] were identified and described as follows:

S. Social introversion: shy, keeps in background on social occasions.
T. Thinking introversion: introspective, reflective, meditative disposition.
D. Depression: often "blue," worries over possible misfortunes.
C. Cycloid tendency: frequent shifts of mood
R. Rhathymia: happy-go-lucky, carefree.
G. General Activity: tendency to engage in overt activity.
A. Ascendance-submission: social leadership or dominance.
M. Masculinity-femininity: similarity of responses to those typical of men or of women.

[18] Three questionnaires were constructed to measure these factors: Guilford Inventory of Factors *STDCR*, Guilford Martin Inventory of Factors *GAMIN*, and Guilford-Martin Personnel Inventory I (*O, Co, Ag*). The thirteen factors are not entirely independent of each other. In fact, Lovell (68), correlating scores on each of the thirteen factors, carried out a factor analysis on these factor correlations and identified four "super-factors": Drive-Restraint, Realism, Emotionality, and Social Adaptability.

I. Inferiority: lack of self-confidence.
N. Nervousness: irritability, jumpiness.
O. Objectivity: viewing self and surroundings objectively, not taking things personally.
Co. Cooperativeness: accepting things and people as they are, tolerant, not fault-finding.
Ag. Agreeableness: not quarrelsome, belligerent, or domineering.

It will be noted that the two investigations found several traits in common. Social introversion, cycloid tendency, and depression were identified in both series of studies.[19] Mosier's "hypersensitivity," moreover, bears considerable resemblance to the Guilfords' "objectivity," albeit expressed in terms of the opposite pole. It is interesting to note, too, that when both studies are considered together, the concept of introversion seems to break down into at least three aspects, viz., social, thinking, and public or "platform" introversion as represented by Mosier's P-factor.

Studies such as the above bring out at least two points. First, the responses of American college students to personality questionnaires tend to be "structured" into a small number of differentiable clusters, rather than being either wholly specific or completely unified and general. Secondly, the labels commonly attached to personality inventories should be regarded with considerable caution. A single test of neuroticism or introversion may measure several independent personality tendencies. Moreover, tests with different labels may measure in large part the same factors, as indicated by the overlap of the factors reported by Mosier and those reported by Guilford in his initial analysis of introversion items. The converse is also likely to be true, viz., tests bearing the same label may measure a different combination of factors.

Mention may also be made of a number of factorial studies of interest inventories such as the well-known Strong Vocational Interest Blank (30, 33, 41, 80). "Interest" factors have been identified which correspond to certain vocational areas, such as "technical science" occupations (e.g., mathematics, chemistry, engineering), social service or welfare work, selling, and financial and business detail work (e.g., accounting, banking, purchasing). These response clusters, moreover, appear to be related to broader aspects of person-

[19] The factor labeled "inferiority," appearing in both lists, was not independently verified, but was taken by Guilford from the Mosier study.

ality, such as values (as measured by the Allport-Vernon Study of Values) and social adjustment. To find that the organization of personality may be significantly related to traditional occupational groupings which have developed within our culture is of considerable interest in connection with the problem of the *origin* of traits.

Some investigators of personality have been engaged in applying factor analysis, not to questionnaire responses, but to *behavior ratings* of both children and adults (18, 25, 26, 27, 28, 70). The most ambitious of these projects is that conducted by R. B. Cattell (28). Beginning with a list chosen to cover all the personality traits which had been named, either in the dictionary [20] or in the psychiatric and psychological literature, Cattell first reduced the list to 171 trait names by combining obvious synonyms. The next step was to obtain ratings for each of these 171 characteristics on 100 subjects of both sexes, all over 25, and varying in occupation from unskilled laborers to artists and business and professional people. Each subject was rated by one person who knew him well, the rating scale containing only two categories for each trait, viz., above average and below average. By correlating these ratings and grouping together all traits which correlated over .45 with each other, 67 clusters were obtained. Through further combination of these groups into "nuclear clusters," the number was reduced to 35. Ratings on 208 men by two independent raters were then obtained for these 35 traits. The men averaged 30 years of age and varied widely in occupation. A factor analysis of the intercorrelations [21] of these 35 traits, followed by oblique rotation of axes, led to what the author terms "the primary source traits of personality." These traits, an even dozen, are given below (28, pp. 475–496).

(1) Cyclothymia *vs.* schizothymia.
(2) General mental capacity *vs.* mental defect.
(3) Emotionally mature, stable character *vs.* demoralized general emotionality.
(4) Dominance and ascendance *vs.* submissiveness.
(5) Surgency [22] *vs.* agitated, melancholic desurgency.
(6) Sensitive, anxious emotionality *vs.* rigid, tough poise.

[20] The dictionary list was based upon the Allport and Odbert compilation (2).
[21] Actually, correlations were computed separately on 13 groups of 16 men each who had been rated by the same raters. These correlations were then averaged via Fisher's *z*-function.
[22] This term refers to placid, realistic cheerfulness and enthusiasm.

(7) Trained, socialized, cultured mind *vs.* boorishness.
(8) Positive character integration *vs.* immature, dependent character.
(9) Charitable, adventurous cyclothymia *vs.* obstructive, withdrawn schizothymia.
(10) Neurasthenia *vs.* vigorous, "obsessional determined" character.
(11) Hypersensitive, infantile, sthenic emotionality *vs.* phlegmatic frustration tolerance.
(12) Surgent cyclothymia *vs.* paranoia.

Some of these traits, such as "general mental capacity" or "a trained mind," overlap with ability variables, but their definition is strongly slanted toward emotional and motivational characteristics. Thus, for example, the terms "deliberate" and "persevering" are included in Cattell's detailed description of "general mental capacity." Cattell has maintained that this list of twelve traits is corroborated by the research of other investigators who used not only behavior ratings but also other methods of trait measurement. Some of the resemblances, however, are not too clearly apparent. In view of the errors to which ratings are known to be subject, the crudeness of the rating scale employed, and other methodological limitations of the present study, much caution must be exercised in generalizing from it. Certainly, data are needed on other and more clearly delineated populations. Perhaps the only observation which can be confidently made at this time is that the search for personality traits has met more obstacles and inconsistencies than that for unitary abilities.

"Common" *versus* "Individual" Traits. Some writers on personality have made a distinction between common and individual traits (cf., e.g., 1). The former refers to the sort of trait identified through factor analysis and other techniques based on standardized tests and on the evaluation of the individual in terms of group norms. The latter, or individual, trait refers to the sort of trait identified by an analysis of the unique experiences of the particular individual. Such a trait, which mirrors the individual's idiosyncratic behavior organization, is observed through clinical procedures and other intensive, prolonged, and relatively qualitative techniques. From one point of view, "type concepts" may be regarded as an attempted compromise between the two extremes of common and individual traits. Such theories imply essentially a pattern of behavior relationships shared by a relatively limited group of people—narrower than the groups to which the common traits of factor analysis are ascribed, but including more than a

single individual. In general, individual traits and type concepts have flourished principally among writers on personality, while common traits have found more support in the classification of intellectual variables.

It should be remembered that, whether found by factor analysis, type studies, or biographical observation of a single individual, a trait is always essentially a *pattern of relationships within the individual's behavior*. The so-called common trait, located by studying a group of persons rather than a single individual, is simply a generalized description of a pattern of behavior relationships shared by a group of persons. Why, then, have such common traits found more ready applicability in the description of intellectual rather than emotional and motivational functions?

The reason is not difficult to find when we consider the greater *uniformity and standardization of experience* in the intellectual than in the emotional and motivational sphere (cf. 9, 10). An obvious illustration of this point is provided by our system of formal education, in which the standardized content of instruction is directed principally toward intellectual rather than emotional development. Even if the schools were to institute a rigidly standardized "personality curriculum" (a rather depressing thought!), we still would not expect the uniformities of organization characteristic of intellectual development, since much of the individual's emotional development occurs through domestic and recreational activities. Not only courses of study, but also occupations and other traditional areas of activity within any one cultural setting, tend to crystallize and structure intellectual development into relatively uniform patterns. Such patterns become more clearly evident the longer the individual has been exposed to these common experiences. The increasing differentiation of abilities with age and education becomes intelligible in these terms, as do the difficulties in identifying common traits among animals.

A further relevant point is the objection raised by some writers (cf., e.g., 1, 81) that test items may have "private meanings," so to speak, for each individual. Discussions of this point have sometimes led to rather cabalistic and obscurantist criticisms of psychological testing. Actually, this objection is simply another way of saying that the same response may not have the same diagnostic or prognostic significance when made by persons of widely varying experiential backgrounds. Since uniformities and standardization of experience in

our culture are more common in the intellectual than in the emotional aspects of behavior, "personality" tests are more subject to such a limitation than are "intelligence" or "aptitude" tests. A further reason for the greater uniformity of intellectual patterns of behavior is found in the degree to which such behavior has been verbalized, as contrasted to emotional responses, which are more largely unverbalized. It may also be relevant to point out that the distinction between intellectual and emotional aspects of behavior is itself culturally determined.

AN EXPERIMENTAL APPROACH TO TRAIT ORGANIZATION

Too often the trait investigator has merely asked: "*What* is the organization of behavior?" or "*What* are the traits into which the individual's behavior repertory groups itself?" rather than asking, "*How* does behavior become organized?" and "*How* do psychological traits develop?" The controversies between exponents of "common traits" and of "individual traits," as well as the apparent inconsistencies in the findings of trait research on different age, educational, or other groups, point up the need for a more direct investigation of the mechanism by which traits develop—the way in which the specific experiential background of different individuals determines the organization of their behavior into more or less unitary and stable traits.

An exploratory study of this question was conducted by Anastasi (6). The principal aim of the investigation was the *experimental alteration* of a factor pattern through a brief, relevant, interpolated experience. Five tests, including vocabulary, memory span for digits, verbal reasoning of the syllogistic type, code multiplication, and pattern analysis, were administered to 200 sixth grade school children. All subjects were then given instruction in the use of special techniques or devices which would facilitate performance on the last three tests only. In its general nature, this instruction resembled that received in the course of school work, as, for example, in the teaching of arithmetic operations, short-cuts of computation, and the like. After a lapse of 13 days, parallel forms of all five tests were administered under exactly the same conditions as in the initial testing. Since the entire experiment was of such short duration, age changes were probably negligible and the influence of other, outside conditions relatively slight.

A comparison of the intercorrelations among the five variables in the initial and final testing showed practically no change in the correlation between the two "non-instruction" tests, viz., vocabulary and memory span. A slight change was found in the correlations between the "instruction" and "non-instruction" tests, and a marked change in the correlations among the three "instruction" tests. Factor pattern analyses revealed marked differences from the initial to the final testing. An examination of the factor loadings in the five tests before and after the instruction suggested that the changes were such as would have been expected from the nature of the interpolated experience.

An interesting parallel in an everyday life situation is provided by a study on the organization of mathematical ability in English school children (76). Wide variations in the correlations among arithmetic, algebra, and geometry test scores were found in different school classes. These variations were shown to be related to such conditions as whether or not the three school subjects were taught by the same teacher, or whether the teaching methods emphasized similarities of technique among these different branches of mathematics.

Also relevant are studies on the effects of practice upon factor patterns. Woodrow (127, 128), for example, found marked changes in the factor loadings of tests following prolonged practice. Nor were these changes a matter of greater reliance upon speed or upon general ability after practice, as might have been expected. Specific changes in the factorial composition of most of the tests occurred in the course of practice, with no evidence for the increasing role of speed or general ability, nor for the presence of a general learning factor.

Such experimental approaches to the development of traits open a way for exploring the mechanism whereby the traits identified in the purely descriptive or cross-sectional studies may have developed. The accumulated effects of education, occupation, and other everyday life activities upon the organization of behavior may be illuminated by a study of the condensed effects of short-range, experimentally controlled experiences.

It has been suggested that in these experiments all that may be changed is the *work method* used by the subject in performing the tests.[23] Such an explanation is certainly plausible, but it should be used consistently. For example, when the test scores of subjects of

[23] Cf. Thurstone, 109, p. 210. For a clear exposition of the role of work methods in individual differences, cf. R. H. Seashore, 85.

different ages, occupations, or educational levels show diverse factor patterns, such differences, too, may be explicable in terms of different methods of work. Moreover, any *uniformity* of factorial organization among members of a given population may be partly the result of commonly acquired methods of work. Factor pattern analyses show only the organization of behavior as it is found in a group of subjects, but do not indicate the origin of such organization.

If we grant that the "traits" identified by factor analysis are simply functional groupings observable within the subject's behavior, then such traits cannot at the same time be conceived as "underlying abilities" which remain unaffected while the subject's method of doing a task and his objectively observable behavior are profoundly altered. Even the common assumption that certain ultimate limits of performance are set by the individual's sensory, neural, and muscular equipment must be modified in the light of the possible variety of work methods. Changing the method of work may in part overcome some of these physical limitations and thus permit the individual to surpass his previously established "capacity level." The whole process of education is, in one sense, a means of changing work methods.

In summary, it would seem that the relationships among the individual's scores on a number of tests *at any one time* may be described in terms of a small number of relatively unitary factors. Under existing cultural conditions, a certain degree of uniformity of factor patterns is found because of general environmental uniformities. Such uniformity of factor patterns is greater in the intellectual than in the emotional aspects of behavior, and probably reflects the influence of traditional educational curricula, vocational classifications, and the like. Thus in the young school child we find a large general factor through all types of activities which are taught in our schools, the so-called higher mental processes. As the child grows older and specialization of function is encouraged, certain culturally determined differentiations appear. "Group factors" are produced in linguistic, mathematical, mechanical, and possibly other functions. These factors, however, are only a mathematical statement or conceptual simplification of the observed relations among concrete responses. And as such they may be expected to shift from time to time in the same subjects or from one population to another because of varying experiences. Such terms as "primary abilities" or "basic traits" are likely to be

quite misleading. They may cause us to forget the real nature of factors.

PART III MAJOR GROUP DIFFERENCES

CHAPTER 16

The Subnormal

IN PART II WE SURVEYED some of the major findings on *individual differences* and attempted to unravel the factors and conditions which produce variation from one person to another. With this background, we may now turn to an examination of certain *groups* into which individuals are commonly classified. Such groupings have been built up through social and cultural traditions and illustrate the general tendency to employ rigid categories and sharp divisions. Thus individuals are popularly classed into the normal and the abnormal, the genius, the feebleminded, the insane, the neurotic. Psychological differences are expected, or at least sought, between the sexes or among nations or "races." Many other groupings can likewise be construed. A person can be classified, for example, in regard to religion, political affiliation, social status, or even place of residence. Psychological differences might be expected between urban and rural populations, or between groups inhabiting regions of different geographical character, such as mountainous or flat, inland or coastal, cold or warm.

These various groupings, like all rigid classifications of individuals, are arbitrary and artificial. In all behavioral traits, people are distributed along a continuous scale and cannot be assigned to distinct categories. When the distributions of any two biologically or culturally differentiated groups, such as the sexes or "racial" and national groups, are compared, the *overlapping* is so large as to render any difference between averages of doubtful practical significance. In such comparisons, the difference between the averages is far smaller than the range of difference within either group. *In the study of individuals, the only proper unit is the individual.* There is no short-cut to the understanding of people, no possibility of learning the behavioral peculiarities of a few broad groups into which any individual could then be conveniently pigeon-holed.

The multiple and complex determination of the individual's behavioral development should in itself make us skeptical regarding any simple systems of characterizing people. Yet it is an all too common practice to expect an individual to be dependable, or shiftless, or dull, or excitable, or poor in mechanics, or to ascribe to him dozens of similar characteristics, simply from the knowledge that such a person is a man or a woman, or that he belongs to a particular "race" or nation.

It is partly to clarify these muddled popular notions that the empirical study of group differences ought to be undertaken. To be sure, a careful examination of the principles underlying individual variation in general should suffice to show the fallacies inherent in many popular claims regarding group differences. But when beliefs are as deep-rooted and emotionally tinged as those governing many group relations, they are not easily dislodged. Direct evidence on the nature of group differences is more convincing than deductions from generally established principles.

From a more theoretical point of view, the analysis of group differences is a valuable adjunct to the investigation of individual differences in general. The existence of culturally diverse groups may be regarded as furnishing a natural experiment in the production of human variability. If psychological differences among groups are investigated with reference to the factors which brought them about, the understanding of individual differences will have been considerably furthered.

In the present analysis of group differences, we shall be more concerned with fundamental concepts and methodological issues than with a cataloguing of results. The latter have little meaning unless critically evaluated. The data on group differences are difficult to interpret and have frequently led to opposite conclusions in the hands of different writers. It is of fundamental importance, therefore, that the special difficulties inherent in group comparisons be clearly realized and that the necessary cautions and controls be applied before making any generalizations. With a clear understanding of the problem, the reader will be in a position to make his own interpretation of any data which he may come across. And he will also be able to guard against hasty generalization and to detect the fallacies in erroneous statements with which he may be confronted. Objective and critical habits of thinking are more urgently needed in the field of group dif-

physiology during the Renaissance. The medical conception of mental abnormalities is still prevalent at present, especially among psychiatrists.

The *psychological study* of the abnormal is of relatively recent date. Its approach to the problem is through a direct study of *behavior*. In some cases, behavioral disorders may have structural concomitants, such as physical diseases, lesions, and malformations. But in the majority of cases no such physical basis has been discovered and it would only obscure the issue to attribute the behavioral manifestations to unknown organic causes. Analysis of the behavioral history and environmental background of the individual, on the other hand, often reveals an adequate explanation for the development of the particular symptoms. Behavior disorders are the special domain of the psychologist and can be studied directly in terms of behavior principles, without vague, hypothetical reference to some other realm or class of phenomena. It should also be noted that abnormality is *specific*. The individual may be quite abnormal in one trait and yet remain close to the norm in other respects. This is true of both intellectual and emotional traits, and follows directly from the organization of behavior traits (cf. Chs. 14 and 15).

Abnormal psychology is an empirical and direct study of behavior deviations. As such it may be regarded as a subdivision of differential psychology. A distinction is now made between feeblemindedness, or intellectual deficiency, and personality disorders. These two categories of behavior deviations will be considered in the sections which follow.

FEEBLEMINDEDNESS

Definition and Levels. Feeblemindedness represents the lower end of the distribution of intelligence. It is characterized by intellectual rather than emotional or personality defect. The term "feeblemindedness" is not used, however, to cover deficiency in *any* ability. Thus an individual may be far below average in music, drawing, or mechanical aptitude, and still be regarded as intellectually normal. *"Feeblemindedness" designates a deficiency only in those abilities which have proved essential for survival in our cultural milieu.*

As was indicated in the preceding chapter, *verbal ability* probably plays the dominant role in our conception of feeblemindedness. Linguistic deficiency has often been explicitly accepted as a criterion of

mental deficiency. Thus Binet and Simon (12) wrote: "An idiot is a person who is not able to communicate with his fellows by means of language. He does not talk at all and does not understand." Similarly, Esquirol (cf. 38, p. 165) distinguished between three levels of feebleminded persons: (a) those making cries only; (b) those using monosyllables; (c) those using short phrases but not elaborate speech. Another classification which is still widely quoted (cf. 38, pp. 165–166) is that which distinguishes between: (a) idiots, who are incapable of spoken language, and are limited to gestures; (b) imbeciles, who are able to understand and employ spoken language; and (c) morons, who are also capable of acquiring written language, but have difficulty with the more complex verbal and abstract concepts.

Feeblemindedness has been described from many points of view. Probably the most common definitions are the *sociological,* or economic, and the *psychometric.* A widely quoted schema of classification, adopted in 1908 by the British Royal Commission on the Feebleminded (14) and still found currently useful (cf. 26, 86), illustrates the *sociological* conception. This classification recognizes three grades of feeblemindedness, characterized as follows:

1. Idiot (low-grade amentia)—"A person so deeply defective from birth or from an early age that he is unable to guard himself against common physical dangers."
2. Imbecile (middle-grade amentia)—"One who, by reason of mental defect existing from birth or from an early age, is incapable of earning his own living, but is capable of guarding himself against common physical dangers."
3. Moron [4] (high-grade amentia)—"One who is capable of earning a living under favorable circumstances, but who is incapable, from mental defect existing from birth or from an early age, (a) of competing on equal terms with his normal fellows, (b) of managing his affairs and himself with ordinary prudence."

The *psychometric* classification is more common among mental testers and permits more quantitative definition. When applied only to adults, the differentiation is often made on the basis of mental age. Thus an adult whose mental age is three years or less is usually regarded as an idiot; between three and seven is the imbecile level;

[4] In England, the term "feebleminded" is reserved for this level of mental deficiency, and "amentia" is used as a general term to cover all degrees of mental deficiency. The term "moron" has been substituted for "feebleminded" in the above definition, in accordance with the more familiar American usage.

morons fall above a mental age of seven but fail to reach the normal adult level. To make the classification applicable to children as well as adults, the limits have been expressed in terms of IQ. Terman's classification [5] (83, p. 79) is probably the most widely employed and has been reproduced below.

Category	*IQ*
Dullness, rarely classifiable as feeblemindedness	80–90
Borderline deficiency, sometimes classifiable as dullness, often as feeblemindedness	70–80
Moron	50–70
Imbecile	20–50
Idiot	below 20

It should be borne in mind that these distinctions are purely arbitrary and are made only for practical convenience. There is no sharp dividing line either between the normal and the feebleminded or between the various "levels" of feeblemindedness. The intellectual differences are of degree only and form a continuous gradation, although the social effects may differ qualitatively. The diagnosis of feeblemindedness, moreover, should never be based solely upon an IQ. The feebleminded individual has been described as subnormal in "personal dependence, self-direction, social responsibility, and self-support" (26, p. 867). A useful adjunct to intelligence tests in arriving at a practical classification of feeblemindedness is the Vineland Social Maturity Scale (24), which measures the individual's "social age" from 0 to 25 years. This scale is a means of evaluating the individual's everyday life behavior in terms of age norms. The subject's emotional balance, health, physique, special skills, and environmental milieu all contribute to the adequacy with which he can cope with everyday life problems; thus they indirectly influence the diagnosis of feeblemindedness in individual cases.

Estimates of the percentage of feebleminded persons in the general population range widely, but the most reliable and comprehensive investigations report frequencies falling between one and two per cent (26). The specific per cents found in different surveys vary with the criterion of feeblemindedness employed—whether tests, social competence, or a combination of the two—as well as with the point at

[5] This classification was originally based upon the 1916 revision of the Stanford-Binet. Subsequent comparisons (64, 71), however, have shown close agreement in the results obtained with the 1937 revision at the lower IQ levels, and the above classification is still considered satisfactory as a rough practical guide.

which the dividing line is set. Geographical locale also makes a considerable difference, the incidence being much greater in some areas than in others. Age, too, will affect the estimate, since the relatively short life expectancy of the feebleminded tends to make the proportion of feebleminded in the population appear greater when only children are included in the survey than when all ages are covered. Similar difficulties are encountered in the attempt to determine the relative frequency of different levels of feeblemindedness. Among institutionalized cases, approximately 10% are idiots, 30% imbeciles, and 60% morons, but the proportion of higher-level cases outside of institutions is probably larger, since such cases are more likely to shift for themselves or to be cared for at home (26).

Varieties and Contributing Factors. The feebleminded have also been classified with respect to variety or clinical type, on the basis of differentiating physical conditions.[6] Among the most familiar of such clinical types is *Mongolism*, named from the oblique, slit-like eyes which produce a superficial resemblance to the Mongolian face. This type can readily be identified by a number of other physical characteristics, such as small, round head; smooth, moist, puffy skin; fissured tongue; and short, stubby fingers. This is one of the most frequent clinical types, constituting from 5% to 10% of the population of most feebleminded institutions. Among the possible causes of Mongolism suggested by different investigations are nutritional, toxic, and endocrine disturbances during uterine life (6, 40). Age of the mother seems to be a factor, the proportion of Mongolians born to mothers over 40 being much greater than the proportion born to younger mothers (34, p. 117).

About equal in frequency to Mongolism is the type of feeblemindedness traceable to *intracranial birth lesions*. As generally used, this category covers not only injuries sustained through instrumental or difficult delivery, but also such conditions as neonatal asphyxia, premature birth, and infectious or toxic factors operating before birth. There has been a growing conviction that many otherwise undifferentiated cases of feeblemindedness may have originated in this fashion (7, 23, 25, 75, 77). Motor disorders of varying degree of severity may be present, as in the familiar "spastic" cases. It is entirely possible, however, for the motor symptoms to develop in an individual

[6] A detailed survey of clinical varieties, their characteristics, and suggested causal theories can be found in Sherman (76) and Tredgold (86).

of normal or superior intelligence.[7] What is more important for our present discussion is that the intellectual defect may occur in an individual without the motor symptoms. In such cases, the cranial injury is not likely to be suspected without an examination into the birth records. The particular combination of symptoms which develops is probably related to the extent and location of the cerebral injury.

Other clinical types of feeblemindedness are relatively infrequent, occurring in less than one per cent of the institutionalized population. The *microcephalic* has an abnormally small, pointed skull, with a characteristic "sugar-loaf" appearance. The *hydrocephalic* has a very large skull and an excessive accumulation of cerebrospinal fluid in the brain. The *cretin* is easily identified by his stunted physique, coarse thick skin, loss of hair, and other physical characteristics. Thyroid deficiency has been clearly identified as a major factor in cretinism. The administration of thyroid extract, if begun early in life, usually effects a considerable improvement in both physical and intellectual condition, although some cases do not respond to this therapy. The causes of microcephaly and hydrocephaly are not so well established, but there is evidence suggesting the role of prenatal factors, including maternal nutrition, toxins, infections, and radiation (72; 76, pp. 145–154).

A relatively rare but clearly identifiable clinical type is *phenylpyruvic amentia* (33, 42, 50, 69). These cases are differentiated by the presence of phenylpyruvic acid in the urine, resulting from a hereditary metabolic disorder. The condition appears to depend upon a single recessive gene (79), and has never been found in a person of normal intelligence (43). It is usually accompanied by motor symptoms and is found in association with a severe grade of mental deficiency.

Mention may also be made of the suggested role of the *Rh factor* in mental deficiency. This is one of the factors determining the "blood groups," which have become familiar to the general public principally in connection with blood transfusions and with the determination of paternity. The Rh factor, discovered in 1939, is peculiar in that it has no natural antibody in human blood, but it may provoke the production of antibodies when introduced into the blood of persons lacking in this factor (rh negatives). It has been estimated that about 15%

[7] Cf., e.g., the interesting biographies of intellectually superior persons with cerebral birth injuries, such as: Hoopes, G. G. *Out of the Running*. Springfield: Thomas, 1939. Pp. 158. Carlson, E. R. *Born That Way*. N. Y.: Day, 1941. Pp. 174.

of the population are rh negative and therefore susceptible to the production of such antibodies. Some important implications of this situation for fetal development have been discovered. First, it should be noted that a certain amount of blood transfusion occurs between mother and child during uterine life. If the mother is rh negative and the child Rh positive, antibodies will be formed in the mother's blood as a result of such a pregnancy. The first-born is not usually affected by this condition, since it takes time for the mother to develop the antibodies. Subsequent offspring, however, if again Rh positive, may develop a severe physical condition which generally proves fatal before birth (29, 58, 78).

It has been suggested that in those cases where Rh incompatibility of mother and child does not result in any observable physical disorders, the effect upon the fetal blood may still be sufficient to interfere with proper brain development and thus indirectly lead to feeblemindedness. Studies of the blood groups of feebleminded children and their mothers have shown that, in a certain percentage of cases not classifiable into any of the known clinical types, the mental deficiency may have resulted from such Rh immunization (20, 21, 79, 86, 90). The per cent of Rh positive children with rh negative mothers in such groups significantly exceeded chance expectation. Subsequent results have tended to qualify these conclusions and to indicate that the operation of the Rh factor in feeblemindedness is probably much less frequent than was suggested by some of the early findings (74, 81). The hypothesis, however, remains a plausible one, at least for a small number of cases, and research along these lines is being actively carried forward.

Those cases not falling into any of the above clinical types are classified as *undifferentiated mental deficiency*. This is by far the largest category, including from one-third to two-thirds of all institutionalized cases (26). The total proportion is probably much greater because such cases, being normal in appearance, are less likely to attract attention and be institutionalized. Other designations for this category are "familial type" and "primary" or "endogenous" mental deficiency, as contrasted to "secondary" or "exogenous." These terms are misleading in their implication that undifferentiated mental deficiency is necessarily hereditary. There is no more reason for associating heredity with undifferentiated mental deficiency than for associating it with the other clinical varieties discussed above. All we

can say positively about undifferentiated mental deficiency is that no physical basis has yet been discovered for it. In some of these cases, specific physical factors may eventually be identified, as illustrated by the recent findings on the Rh factor and on intracranial injuries. For the rest, it is possible that the feeblemindedness is not associated with any structural deficiency but only with experiential factors.

The fact that undifferentiated mental deficiency tends to run in families (hence the designation "familial type") may, of course, be interpreted as evidence for environment just as well as for heredity. In this connection, it is interesting to note that the few studies which report improvement in intellectual level as a result of special training have found that it is the undifferentiated type that responds most readily to such training (48, 49). It is also relevant to observe that undifferentiated aments more often come from homes of lower socio-economic level, while the specific clinical types show a more random distribution of home background (13, 35). The latter occur with greater frequency than the former in families which are normal or superior in intellectual and socio-economic level. Inferior home environment may be a factor in the intellectual retardation of at least some of the cases in the undifferentiated group.

Certainly the term "undifferentiated" is more precisely descriptive of our knowledge regarding this type of mental deficiency than are the other suggested designations—"unknown" would probably be a more candid characterization. To assume a hereditary basis for just those cases in which no structural deficiencies have yet been demonstrated seems to suggest that feeblemindedness is itself a chemical substance which can be transmitted by the genes! Unless some structural deficiency is demonstrated, *what* is there for these cases to inherit? The evidence for hereditary contributions seems, in fact, to be much clearer in the case of the so-called secondary forms of feeblemindedness. Glandular and metabolic conditions of the mother, blood groupings, and even maternal body formation which might increase the chances of a difficult birth undoubtedly have a hereditary basis. From one point of view, of course, it may be argued that the feeblemindedness in such cases is only an indirect result of the hereditary condition. But this only serves to point up the artificiality of the heredity-environment distinction, especially as applied to behavior. Of more practical significance is the distinction proposed in Chapter 4 between structurally and functionally determined conditions. The specific clinical

types of feeblemindedness discussed above are all structurally determined and as such would be relatively uninfluenced by training. In such cases, the structural deficiency interferes with the acquisition of normal behavior. Many of the "undifferentiated" cases, on the other hand, may be functionally determined and therefore much more responsive to training.

Intellectual and Physical Status. It has been repeatedly demonstrated that the feebleminded are not equally deficient in all functions and that the degree of their inferiority increases as we go from simple sensory and motor tasks to complex intellectual processes, especially those dealing with symbols (17, 38, 62, 66). In a pioneer study, Norsworthy (66) administered a series of tests, including comparison of weights, cancellation, memory, and word association, to 157 institutionalized defectives. In the comparison of weights, 28% of the feebleminded reached or excelled the performance of the lowest quarter of normal subjects.[8] In the cancellation tests, the per cent of feebleminded reaching or exceeding the lowest quartile of the normal group ranged from 14 to 18, in the memory tests from 18 to 19, and in two of the association tests from 16 to 17. In the two remaining association tests, which involved the naming of opposites, only about 1% of the feebleminded reached or exceeded the lowest quartile of the normal These findings were corroborated by Merrill (62) in a comparison of the performance of mentally deficient and normal children on the separate tests of the Stanford-Binet. The achievement of mental defectives in different school subjects shows a similar hierarchy (17, 62). The feebleminded as a group are most deficient in verbal subjects, such as reading comprehension, less deficient in arithmetic computation, and closest to the norm in drawing and shop work.

This does not mean, however, that such a hierarchy of deficiency necessarily exists within the individual feebleminded person. The same result might follow if there were *more* feebleminded persons deficient in verbal ability, fewer deficient in arithmetic ability, and fewest in mechanical or sensori-motor aptitudes. The relationship between such group averages depends not only upon the relative amount of inferiority displayed by each individual, but also upon the *number* of persons who are inferior. Since verbal aptitude plays such a large part in the criterion of feeblemindedness, almost all persons in a feeble-

[8] If the two groups were equal, 75% of the feebleminded group would reach or exceed the lowest quarter of the normal group

minded group will be definitely below normal in this trait. This consistent inferiority will of course produce a very low group average in verbal traits.

In numerical aptitude, many will be below average, some may be normal, and a few even superior. The slight positive correlation between performance on numerical and verbal tests, as well as the fact that numerical tests are frequently included in scales of "general intelligence," would lead us to expect the *majority*, but not all, feebleminded persons to be below the norms in numerical aptitude. This would result in a group average higher than that in verbal traits, but still considerably below normal.

In tasks involving sensori-motor skills or aptitude in mechanics, music, or pictorial art, we should expect the feebleminded distribution to approximate even more closely that of a normal group, since these traits show very low correlations with verbal ability or intelligence test performance. The majority of the feebleminded would be nearly normal in these functions, only a small number inferior, and a few superior. As a result, the status of the group as a whole would be only slightly below normal. Thus it is apparent that the hierarchy of deficiency usually found in feebleminded groups may result from the culturally imposed criterion of feeblemindedness and from the organization of abilities.

Many observers have called attention to the *rigidity* and *stereotypy* characteristic of the behavior of mental defectives.[9] Institutionalized morons, for example, will often carry out routine tasks with unswerving precision and with no signs of boredom. Such persons seem well qualified for monotonous, repetitive tasks. It is, of course, well known that monotony is a function of the nature of the task, the distracting stimuli, and the characteristics of the worker. The individual with relatively few competing interests and limited ability will find a repetitive task more congenial and satisfying than a task involving many shifts and readjustments. Despite the evidence of "rigidity" in the behavior of many feebleminded persons, it should be noted that the feebleminded are capable of considerable *improvement through learning*. Thus in an experiment with simple sensori motor and perceptual tasks, feebleminded adolescents improved with practice about as

[9] Kounin (53), for example, reports some suggestive data obtained with tests of rigidity on older and younger feebleminded persons and on normal children, all groups being equated in mental age.

rapidly as normal children of the same mental age (89). Such a finding suggests that all but the lowest-level cases can certainly profit from training, if the tasks and methods of instruction are suited to their mental age level.

In *general health, susceptibility to disease,* and *physical development,* the institutionalized feebleminded *as a group* are below normal. Data on this question have already been discussed in Chapter 12. It will be recalled that such comparisons must be accepted with caution because of the inclusion of physically defective clinical types, the low socio-economic background of most cases, and the selective factors in institutionalization. It is certainly not difficult to find *individuals* among the higher grade undifferentiated mental defectives who are sturdy, healthy, and good-looking by normal standards.

Outlook for Social Adjustment. Idiots and most imbeciles obviously require either institutional or home care. Within the much larger group at the moron level, however, a considerable proportion of individuals are "on their own." It is this group that has been a source of concern as a potentially serious social problem.[10] Recent follow-ups have shown that, with a certain minimum of training and supervision, the outlook for such mental defectives is more favorable than was formerly supposed (76). Persons who have acquired good work habits and mastered a simple skill in a feebleminded institution have sometimes succeeded in their jobs as well as or better than normal persons with a long work history.

This should not be surprising when we consider the number of jobs in our society which do not demand a high intellectual level. One survey of over 2000 jobs, conducted with special reference to the vocational guidance of the mentally defective, revealed 19 types of occupations for which a minimum mental age of only 6 years was required (16). A number of other kinds of work could be successfully performed by persons with mental ages of 7 to 11 (16). In another, more extensive survey of 18 industries providing 2216 specific occupations, 47.1% of the jobs required no education beyond the ability to speak, read, and write simple English (5). A total of 67% called for no education beyond elementary school graduation. Even in these cases, it is doubtful whether some of the more abstract con-

[10] Cf., e.g., the dramatized presentation of this problem in F. R. Wembridge, *Life Among the Low-Brows.* Boston: Houghton Mifflin, 1931. Pp. 310.

tent of the elementary school curriculum, which might be beyond the grasp of the average moron, was really necessary for job success.

Among individuals *paroled or discharged* from feebleminded institutions, the proportion who adjust satisfactorily is undoubtedly large enough to justify such a parole practice. Specific estimates are difficult to summarize because of varying standards of successful adjustment applied by different investigators. The estimates of successful social adjustment, without reference to earning capacity, range from about 50% to 72% of those paroled (22, 68, 82). The proportion is much smaller when adequate vocational adjustment is considered. One estimate reached from a consideration of several available surveys (76) sets the proportion of those adjusting adequately on an economic and social basis without supervision at only about 5%. An additional 20% can be self-supporting with supervision; much larger percentages can do some productive work but not to the extent of being self-supporting; and still others, although unable to hold any job, can remain with their families without creating social problems. These estimates are rather conservative, and it should be remembered that they are based on surveys which were conducted largely during a period of economic depression, when jobs were hard to find. In contrast, it is interesting to note the findings of a survey of 177 young people [11] paroled in 1941–42 from a training school for mental defectives (37). Within this group, 88% were employed, many above the level of unskilled labor. Most were earning from $40 to $60 a week, had held their jobs over three months at the time of the survey, and had secured them without the assistance of family, friends, or social agencies.

Follow-ups of children who have been diagnosed as mentally defective and placed in *special classes* in the public school system show, in general, a larger proportion of successful adjustment. This is to be expected, not only because such groups are likely to include children of higher intellectual levels, but also because commitment to an institution is often an indication of poor social adjustment coupled with mental defect. In one follow-up (27), 122 out of 166 Baltimore school children who had been diagnosed as mentally defective

[11] A total of 211 individuals had been paroled, of whom 6 had been re-institutionalized or were described as unemployable at the time of the survey, and 28 could not be located. The present analysis was based on the remaining 177 cases.

were located 17 years later. Over 75% of this group had never required any support from a social agency. The large majority had had no brushes with the law, although the number with court records was about three times as large as that in a normal group from the same school district. In interpreting the latter finding, it should be remembered that a large number of mental defectives come from squalid, depressing, and unhappy homes, surroundings which frequently lead even the intellectually normal to jail.

In a later and more thorough investigation (2), a similar group of "opportunity class" children in Nebraska were investigated when all were between the ages of 21 and 34. Of the original group of 206, 196 were located and compared with a high-normal group whose IQ's ranged from 100 to 120. Slightly less than 7% of the original mentally defective group were in institutions for the feebleminded at the time of the follow-up. Educationally, the subnormal group had completed an average of 4 or 5 grades, in contrast to the 12- to 13-grade average of the "normal," control group. As in the other survey, the majority had no court records, although the proportion with such records exceeded that in the control group: 25% *vs.* 4% for juvenile court, and 18% *vs.* 6% for police court. The proportion of the mental defectives who had held relatively permanent jobs was 39%, as compared with a proportion in excess of 90% for the normal, control group. Among the subnormal, however, 83% had been partially self-supporting for varying periods of time. The proportion of girls who had married was about equal in the two groups, although the subnormal girls tended to marry earlier and have more children. Among the boys, the percentage who had married was much smaller for the subnormal, probably because of economic reasons. In a later follow-up on an intermediate, "dull" group, more delinquency and poorer social adjustment were found than in the mentally defective group (3).

Several additional points should be considered in evaluating the findings of such surveys. First, a large number of these "special class" children come from inferior homes in which the parents are unable to provide adequate direction. With proper supervision, many more cases could probably adjust socially as well as vocationally. Secondly, job turnover is very common among such groups and is taken as an index of poor vocational adjustment. In at least some of the cases, such turnover could probably be avoided by better vocational

counseling. Thirdly, the fact that the more *recent* surveys tend to show more favorable outcomes may result in part from the improved training and guidance facilities for such groups, both in institutions and in the school system. In this connection, mention may also be made of the suggestive findings reported by Schmidt (73) in the case of a specially designed educational program (cf. Ch. 8).

PERSONALITY DISORDERS

Psychoses. "Insanity," more technically known as "psychosis," represents a pronounced maladaptive personality deviation. In such a condition, the individual, although often intellectually normal or even superior, is unable to make a satisfactory adjustment because of serious personality disorders. Thus he may have delusions of persecution which make him suspect all with whom he comes into contact of plotting to poison him, or delusions of grandeur in which he believes

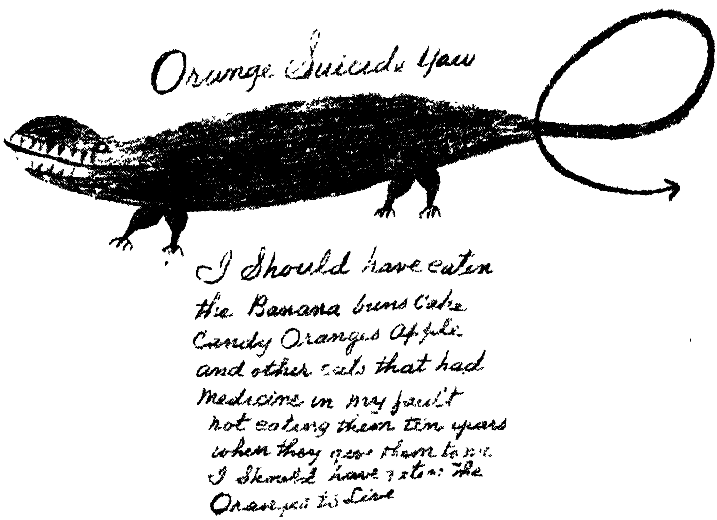


Fig. 85. Delusional Animal Drawn by a Paranoid Patient. (From the authors' collection.)

himself to be Napoleon or some other favorite character. Such symptoms are described as *paranoid.* Drawings by two paranoid patients are reproduced in Figures 85 and 86. The first of these shows the "Orange Suicide Yaw," an imaginary animal which the patient believed to be lodged in his stomach and to which he attributed all his difficulties. The animal is entirely black, with the exception of a bright orange tongue. Figure 86 is a typical astronomical drawing by a patient who had the delusion that he alone had received secret knowledge which explained the motion of the earth.[12]

The individual may withdraw so far into his own fantasy-life that he loses all contact with his fellow-beings and with occurrences about him, as in *schizophrenia.* Also characteristic of schizophrenics are such symptoms as hallucinations; disorganization of thinking and doing; strange, bizarre activity; and odd, stereotyped mannerisms and posturing. It has been estimated that schizophrenics constitute nearly half of the total resident mental hospital population (56, p. 43). The incidence in the general population has been placed at approximately 0.85%. Another frequent category, occurring in about 0.44% of the general population, is the *manic-depressive* group of psychoses, characterized by recurrent periods of extreme depression and excitement or overactivity.[13]

Such psychotic conditions are not to be confused with feeblemindedness. Psychotics are recruited from all intellectual levels, the majority falling within the normal range of intelligence. Instances are not unknown among the intellectually gifted. Psychotic conditions may likewise occur among feebleminded persons, although for some psychoses a certain minimum complexity of intellectual development seems to be required. Certain psychoses, such as schizophrenia, often lead to intellectual deterioration, but there are others in which the patient may suffer no impairment of abilities.

As in the case of feeblemindedness, there is no sharp dividing line between "insanity" and normality (63). Distinctions are made for practical purposes of confinement, treatment, and similar reasons, but close examination reveals a continuous, unbroken gradation from the thoroughly well-adjusted person to the conspicuously insane.

[12] For a brief report of the research project in which these drawings were collected, cf., 1.

[13] For a classification and description of psychotic disorders, cf. any recent standard textbook on abnormal psychology, such as 18, 19, 56, 63.

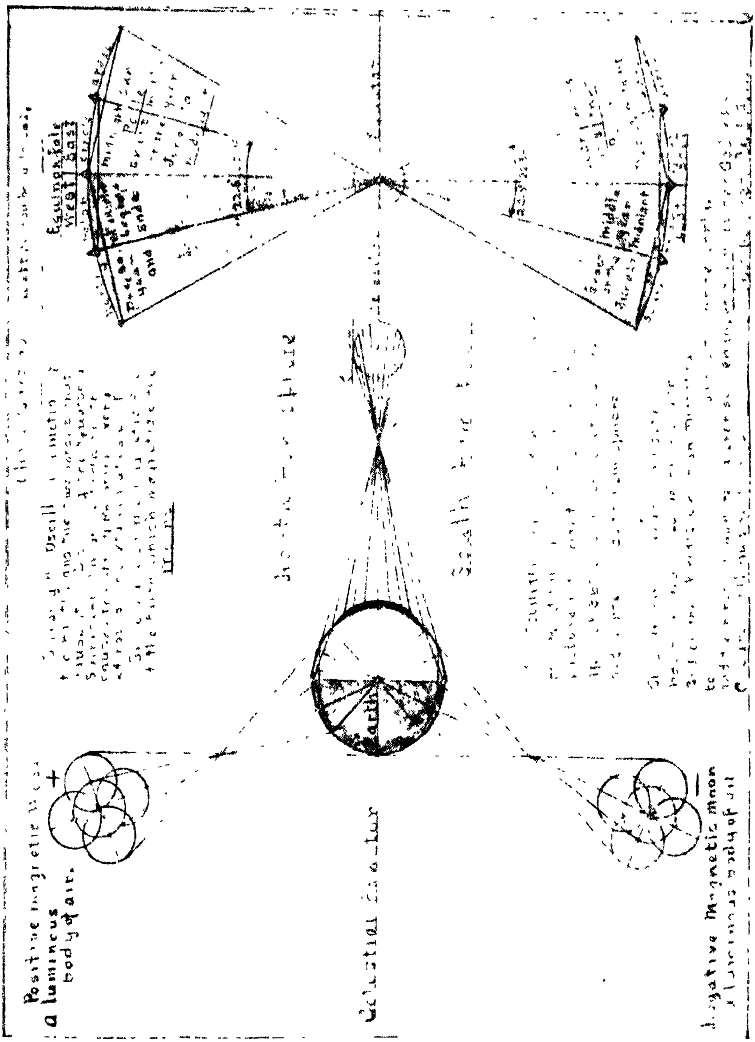


Fig. 86. "Astronomical Diagram" by a Paranoid Patient. (From the authors' collection.)

Psychotic symptoms differ in degree from the behavioral peculiarities of the normal individual. From the blissful optimist who trusts implicitly whomever he meets, to the paranoiac who believes that the stranger who accidentally brushes against him is plotting his demise, there are all degrees of "suspiciousness." The same may be said of other characteristics of the insane. A good example of this is the familiar case of the student who, upon reading a manual of psychiatry or attending a course in abnormal psychology, believes himself to be afflicted with each form of psychosis in turn. Most of us can discover in ourselves at least one characteristic of many types of insanity, in mild form. It is not normal, in the statistical sense, to be entirely free from all such slight peculiarities.

An important distinction from the viewpoint of the psychologist is that between *organic* and *functional* disorders. Briefly, organic disturbances are those which can be definitely correlated with a structural deficiency. In functional disorders, on the other hand, there seems to be only a faulty operation or deficient action of apparently normal structures.[14] Thus paresis has been definitely traced to the influence of syphilitic infection upon the nervous system; a group of psychoses have been shown to develop from excessive use of alcohol or drugs; injuries, or lesions, in certain parts of the brain or lower nerve centers lead to characteristic behavioral deficiencies. There remain, however, a large number of psychotic conditions for which no physical basis has been discovered. These constitute the most common types of psychoses, including schizophrenia and the other psychoses discussed in the early part of the present section.

A few psychologists and psychiatrists are of the opinion that the physical bases of such behavioral disorders are only undiscovered and that ultimately all will be adequately explained in structural terms. There is a growing conviction, however, that these disorders may be purely functional in their origin, involving no structural impairment. If this be the case, we should seek the causes of such psychotic conditions in the mechanisms of learning and in the environmental conditions which have surrounded the individual throughout his lifetime.[15] The question is still a controversial one, to which a conclusive answer cannot yet be given. The fact that no differen-

[14] This is essentially the same distinction as that between structural and functional determination made in the section on feeblemindedness and in Chapter 4.

[15] For a vivid analysis of the way in which such disorders might be built up through learning, cf. 87.

tiating physical characteristics have been discovered cannot, however, be attributed to a dearth of data. Investigators have certainly *tried* to find such differences—in the brain or other parts of the nervous system, in the endocrines, in the chemical composition of the blood, and in innumerable other physical factors—but the findings so far have been negative (18).

One finding which some psychologists have regarded as strong evidence for the heredity of psychoses is that such conditions tend to *run in families*. The most extensive surveys on this question are those conducted by Kallman (44, 45, 46, 47). In the first of these surveys, Kallman obtained data on the relatives of 1047 schizophrenic patients admitted to the Herzberge Hospital in Berlin between 1893 and 1902. One of the results of this study concerns the incidence of schizophrenia among parents and children. Within the entire group of families in which both parents were schizophrenic, 68% of all the children were also schizophrenic. When one parent was schizophrenic and the other schizoid,[16] the proportion of schizophrenic children was 24%; when one parent was schizophrenic or suspected of schizophrenia and the other normal, 15%; and when both parents were themselves normal but had schizophrenia somewhere in their ancestry, 9%. Among unselected families, the proportion of schizophrenic children would of course be about 0.85%, since that is the incidence of the condition in the general population. These results are of considerable interest and value in themselves, but they do not, of course, provide any more conclusive evidence for either heredity or environment than do other similar data on family resemblances discussed in Chapter 10.

In the same study, further data were presented on the half-brothers and half-sisters of schizophrenic patients (44). These figures showed that when the common parent was schizophrenic, the incidence of schizophrenia among the half-siblings of the schizophrenic patient was much higher than when the common parent was normal (24% *vs.* 2%). Such data seem to provide a sort of environmental control, since the half-siblings had been living in the same family. There are difficulties, however, in these comparisons. First, the number of half-siblings in the two categories was fairly small. Secondly, the comparison is not quite so clear cut as might at first appear, since within the group with a common *normal* parent there would obviously be more families in which one or both parents were normal than there

[16] A normal personality showing mild schizophrenic behavior tendencies.

would in the group with a common *schizophrenic* parent. In other words, the total family environment of the group with a common normal parent may also have been more favorable and normal. Thirdly, the contact between parent and own-child may have been psychologically closer than that between parent and step-child. Finally, we cannot ignore the possible effects upon the half-siblings of their knowledge concerning normal or abnormal parentage; nor can we ignore the possible influence of such knowledge upon the reactions of family and associates toward the siblings.

Additional analyses were conducted on twins (46, 47). In identical twin pairs in which one twin had schizophrenia, 82% of the co-twins were suffering from the same condition. In non-identical pairs, the per cent was 12.5, and among siblings 11.5. In interpreting these figures, all the factors which tend to make the environment of identical twins more alike than that of fraternal twins or siblings must be taken into account (cf. Ch. 11). In a later survey (45) conducted by Kallman in mental hospitals in New York State, 691 twin pairs were located in which at least one member of each pair was schizophrenic. The relative frequency of schizophrenia among the co-twins and other relatives of the schizophrenic twins closely corroborated the earlier findings. Within the total group of 174 identical twin pairs, 85.8% were concordant, i.e., both members were diagnosed as schizophrenic. When 59 pairs of identical twins who had been separated for 5 years or more prior to the onset of the psychosis were compared with the remaining 115 non-separated pairs, the incidence of concordance was 77.6% for the former and 91.5% for the latter.

Several case reports of identical and non-identical twins are also given by Kallman. Among them is the account of a single pair of separated identicals, both of whom developed schizophrenia at the age of 24 although they had been reared apart since infancy (44, pp. 207–209). In contrast to this case is that of a pair of fraternal twins, one of whom developed an acute and deteriorating form of schizophrenia, while the other remained normal, despite the fact that both had been exposed to the same extremely unfavorable environment (47). The normal twin in this pair was also physically stronger and more mature, a finding which Kallman reports for the majority of twin pairs with only one schizophrenic member. This difference in physique could, of course, operate as an environmental factor in

the subsequent development of the psychotic condition (cf. Chs. 11 and 12). On the whole, these data are no more conclusive for the question of heredity than most investigations on family resemblances and differences. Moreover, the fact remains that if there is a hereditary basis to schizophrenia, we should be able to show *what* it is that is inherited.

Similar studies on twins and other family relationships have been conducted in the case of other psychotic disorders, such as manic-depressive psychoses (cf. 56). In general, their findings are similar to those obtained with schizophrenics, and their interpretations are subject to the same limitations which have been discussed above. Psychoses do "run in families," but more than that we cannot justifiably conclude.

Neuroses. The neuroses, also known as "psychoneuroses," may be regarded as milder forms of personality disorder than the psychoses. They are also more generally considered to be functional or "psychogenic" in origin than are the latter. In their specific manifestations, they bridge the gap between the slightly maladjusted individual on the one hand and the distinctly psychotic on the other. On the basis of the normal distribution of behavioral characteristics, we should expect neurotics to be more numerous than psychotics, since they are nearer the center of the curve. This seems to be quite clearly borne out by actual statistics. Precise data on the incidence of neuroses are difficult to obtain, however, because of inconsistencies in diagnostic criteria or standards and because most neurotics are not institutionalized.

For convenience of classification, neurotic symptoms have been grouped into a few major clinical pictures. It should be borne in mind, however, that individual neurotics may be quite unique and that pure "textbook cases" are rare—probably even rarer among neurotics than among psychotics. Among the most commonly employed categories are psychasthenia, hysteria, neurasthenia, and anxiety neuroses [17] (18, 19, 56, 61, 63). Typical symptoms of *psychasthenia* include obsessions (persistently recurring thoughts), phobias (unwarranted fears), and compulsions (e.g., continuous and needless washing of hands).

Hysteria is generally characterized by loss or impairment of bodily

[17] This terminology is used here because it is very common and likely to be encountered by the student. A number of psychologists have been taking steps toward a much needed reclassification of behavior disorders, but the classification is still in a state of flux (cf., e.g., Cameron, 19).

function, but in some cases may involve such symptoms as amnesias or "trance states." Among the most typical hysteric symptoms are loss of movement in some part of the body and sensory impairments, such as deafness, blindness, or skin anaesthesias. These physical symptoms of the hysteric, however, are entirely functional in origin, and can readily be distinguished from organically caused paralyses and loss of sensation. Hysterical symptoms are often anatomically impossible, as in the "stocking" or "glove" anaesthesias, which represent bodily regions that are unitary only in popular thought but do not correspond to the known distribution of nerves. Sometimes such symptoms are intermittent and occasional, and may be manifested only in the presence of certain individuals or in a particular locality. Another distinguishing feature of hysterical symptoms is their susceptibility to a wide variety of "cures" based upon suggestion. Undergoing practically any sort of experience in whose efficacy the patient believes has been known to produce many sudden and startling cures. The history of such cases usually reveals a certain obscurity of diagnosis: physicians were baffled, the case was declared hopeless or a mystery. Needless to say, such a case history adds to the glamour of the "cure." What this really shows, of course, is the functional nature of the disorder which naturally defied organic diagnosis and treatment. The popular prestige of many unscrupulous charlatans is built upon their success in "curing" the apparently physical disabilities of certain hysterical patients.

Neurasthenia represents a condition of intense mental and physical fatigue induced by prolonged emotional maladjustment. The individual feels "fagged out" and "a general wreck," often suffering from insomnia, lassitude, loss of appetite, and physical complaints of a hypochondriacal nature. *Anxiety neuroses* are characterized by attacks of intense fear, together with the physical symptoms that accompany this emotional state. An example is "combat fatigue," which has also been classified as a traumatic neurosis, i.e., a neurosis resulting from a severely disturbing or injurious experience. Recurrent nightmares and an exaggerated startle response to such mild stimuli as the slamming of a door are characteristic symptoms of this well-known war neurosis.

Neurotic symptoms are no less "real" because they are functional. The subject may suffer just as acutely and be as seriously handicapped as if he had a definite structural deficiency. Similarly, such

disturbances cannot be overcome merely by voluntary effort. Nor should they be confused with *malingering*. The subject himself is undergoing as vivid an experience as if he had an organic disorder and he may be completely unaware of the fact that his symptoms have no structural basis.

Like psychoses, neuroses show little relation to intelligence and probably occur at all intellectual levels except the lowest (39, p. 975). Biographies of professionally eminent persons who have developed acute neurotic conditions bear witness to the fact that neuroses are not incompatible with very high intelligence.[18] A few studies with personality inventories do suggest that the more intelligent individuals are less likely to give certain neurotic responses (15, 60), but this might result from a more sophisticated response to the test items on the part of the brighter individuals. Moreover, differences in socio-economic level, which are correlated with intellectual status, may account for any observed differences in neurotic tendencies. The latter explanation may also apply to possible differences in the relative frequency of each type of neurosis at various intellectual levels (39, p. 975). Individuals who, by virtue of differences in education, occupation, and the like, are exposed to different situations are likely to differ in the nature of symptoms which they develop.

It has frequently been said that neuroses are the result of the stress and strain of modern living, especially in the more hectic urban and metropolitan centers. This may be partly true, but the statistics usually quoted in support of such statements should be examined with considerable care. With improvements in methods of diagnosis and facilities for treatment, many more neurotics are *recognized as such* today than were recognized twenty or forty years ago. Many neurotics are not sufficiently maladjusted to attract much attention or to demand urgent treatment. Heretofore, such individuals may have gone their unhappy way, probably unpopular or disliked among their associates, but bearing their difficulties unlabeled and unrecorded. The more highly developed the methods of diagnosis, the milder will be the disorders which can be detected and the more numerous the individuals classified as neurotic. The same argument applies to the statistics for urban centers, where psychiatric facilities are much better than in rural areas (57).

[18] Cf., e.g., Leonard, W. E. *The Locomotive God.* N. Y.: Appleton Century, 1927. Pp. 434.

ABNORMALITY IN DIFFERENT CULTURES

Varieties of Normality. Psychologically, all behavior follows normally from its antecedent conditions—there is no essential distinction between the mechanisms or psychological principles of normal and abnormal behavior. Abnormality is the normal consequence of certain stimulating conditions and structural characteristics. Behavior is abnormal only in the sense that it deviates from a norm. This norm is determined by the specific conditions of life within a given group. Thus it follows that behavior which is considered abnormal in one culture may be normal in another.

Cultural standards enter into the definition of normality in at least two ways (31). First, the *position of the norm* and the line of demarcation between normality and abnormality may differ from one group to another. As a result, any given behavioral manifestation may occupy a very different place in different distributions of behavior. To take an illustration from physical traits, if we ask whether a man is tall or short, we may obtain very different answers when different groups are employed as standards. The same individual might be abnormally tall when referred to the distribution of height in the Japanese and very short when referred to the Scandinavian distribution. Similarly, in certain groups violent displays of emotion are the rule and stolidity would be abnormal. In others, the reverse is true. The range of variation over which normal behavior may occur can also differ. Thus two cultures having the same norm may differ in the degree of deviation from this norm which is possible without maladjustment. In one, rigid adherence to a narrowly defined behavioral norm may be required, either because of tradition or because of the exigencies of the physical environment. In another, wider latitude and larger individual differences may be acceptable as "normal."

In the second place, culturally established standards may determine *which end of the distribution is superior and which subnormal.* Comparative anthropology provides many examples of behavioral deviations which are regarded as unadaptive, pathological, insane, or mentally deficient in one culture and are admired or revered in another. Such behavior may be abnormal in both cases, in the statistical sense, but its social evaluation and practical value in the different cultures place it at opposite ends of the scale. This point was clearly expressed by Benedict (8), who wrote:

> . . . it is probable that about the same range of individual temperaments are found in any group, but the group has already made its cultural choice of those human endowments and peculiarities it will put to use . . . the misfit is the person whose disposition is not capitalized by his culture. . . . It is clear that there is not possible any generalized description of "the" deviant—he is the representative of that are of human capacities that is not capitalized in his culture (p. 24).

The same point of view was further elaborated in a later article (9) by Benedict as follows:

> One of these problems relates to the customary normal-abnormal categories and our conclusions regarding them. In how far are such categories culturally determined, or in how far can we with assurance regard them as absolute? In how far can we regard inability to function socially as diagnostic of abnormality, or in how far is it necessary to regard this as a function of the culture?
>
> As a matter of fact, one of the most striking facts that emerge from a study of widely varying cultures is the ease with which our abnormals function in other cultures. It does not matter what kind of "abnormality" we choose for illustration, those which indicate extreme instability, or those which are more in the nature of character traits like sadism or delusions of grandeur or of persecution, there are well described cultures in which these abnormals function at ease and with honor, and apparently without danger or difficulty to the society (p. 60).

Among the natives of Dobu, an island in Melanesia, fear, suspicion, and mutual distrust characterize the attitudes of the entire group (32). They take constant precautions against being poisoned or having their property removed by sorcery or trickery. Within our culture such behavior would be described as paranoid, but it represents a normal adjustment to the Dobuan culture. Illustrations can easily be multiplied (10, 51, 52, 54, 56). The cataleptic seizures constituting an important part of the behavior of the Siberian shaman and the homosexual practices common in many American Indian and Siberian communities represent other illustrations. Trance states are a normal part of the behavior repertory of certain American Indian groups, and it is the individual who is unable to experience the trance who is the deviant.[19] Epileptic seizures, excessive daydreaming, and withdrawal characterize the superior deviant in certain cultures, rather

[19] Cf., e.g., the interesting biography of such a deviant recorded in Radin, P., ed. *Crashing Thunder: the Autobiography of an American Indian.* N. Y.: Appleton-Century-Crofts, 1926. Pp. 202.

than being a source of maladjustment. To be sure, the "significance" of such behavior for the individual differs from that in our culture. But this is just what we mean by saying it is normal in one culture and abnormal in another. In one case, the individual is behaving in a manner which is sanctioned and overtly encouraged by his culture; he is conforming to the accepted and institutionalized pattern. In the other case, he is not.[20]

Varieties of Abnormality. All cultures have their deviants and their maladjustments. But the form which such maladjustments take may vary widely with the cultural setting (4, 10, 51, 52, 56). In the *windigo psychosis* among the Ojibwa Indians, the individual believes he has been transformed into a windigo, a mythical cannibalistic giant made of ice (55). The condition usually begins with a state of depression and often develops into violence and compulsive cannibalism, in which the individual may kill and eat the members of his own family. Other familiar examples include *arctic hysteria*, found in northern Siberia, in which the individual shows a high degree of suggestibility and compulsively imitates the words and actions of those in his vicinity. A similar condition found among the people of Malay is known as *latah* (70). Also characteristic of the Malayan culture is *amok*. The person who "runs amok" attacks in a blind rage everyone he meets, frequently injuring or killing many before he is stopped. The influence of cultural factors upon the specific nature of deviant or maladjusted behavior was also illustrated by a survey of the neuroses observed among native African troops during World War II (65). The relative frequency of certain types of symptoms, such as phobias and hysterical symptoms of a motor or sensory nature, and the almost complete absence of other conditions, such as anxiety states, could best be understood in terms of the particular tribal beliefs and traditions.

We need not go to "primitive" peoples for illustrations, but can find them in our own cultural history (cf. 52). The dancing manias which swept over whole villages in the Middle Ages are a form of neurotic behavior having no direct counterpart today. Many of the manifestations of witchcraft provide further illustrations. The trance states,

[20] Wegrocki (88) has argued against the cultural and statistical concept of abnormality on the grounds that the same behavior may be indicative of severe maladjustment in one culture and of good adjustment in another. Far from being a criticism of the cultural concept of abnormality, this follows directly from it, as indicated in the above discussion.

the hysterical insensitivities such as the "devil's claw"—an insensitive spot on the skin often used as "evidence" in witchcraft trials—were all part of a clinical picture which fitted into the culture of its time. A further example of such "fashions in abnormality" is the delicate, languishing type of illness of unknown origin which was so common among Victorian gentlewomen.

ABNORMALITY IN INFRAHUMAN ORGANISMS

Other species have their deviants too. Mental deficiency, as well as "unadaptive" behavior which can be characterized as psychotic or neurotic, has been noted in many animal forms. Homosexuality has been observed or experimentally induced among doves, pigeons, guinea pigs, white rats, and monkeys (cf. 36, 41). Several investigators working with monkeys have reported instances of other types of abnormal behavior, such as habit residuals, temper tantrums, infantile reversions, and various forms of sexual perversions (30, 36, 84). These constitute abnormalities in the sense that they differ conspicuously from the usual behavior of the species. Whenever the etiology of such abnormal behavior could be definitely traced, experiential or environmental factors were found to play a predominant part in its development (30, 36, 41, 84).

In the course of his conditioning experiments with dogs, Pavlov (67) observed several instances of distinctly neurotic behavior. Such behavior appeared when the animal was required to make too fine a sensory discrimination, or to set up too many conditioned reactions within a short time, or to establish a conditioned reaction when the two stimuli were separated by too long an interval. The neurotic behavior included violent emotional display and loss of previously established discriminations. These early observations of Pavlov stimulated a number of investigators to explore the problem further, and the production of "*experimental neuroses*" in animals has become a common research procedure (59). Several modifications of the Pavlovian conditioning technique have been employed, as well as a number of conflict-producing situations. The neurotic behavior observed varies from sleepiness, inertness, and rigid immobility to hypersensitivity and overactivity, sometimes reaching manic excitement. Symptoms resembling chronic anxiety, phobias, infantile regression, compulsions, and hallucinatory phenomena have been described.

CHAPTER

17

Genius

FROM EARLIEST TIMES, man must have been aware of the genius in his midst. In order to be recognized as a genius, the individual must display an unusual degree of the talents demanded by his culture. Since only the extreme deviates attract notice, they seem by the very rarity of their attainments to stand off from the rest of mankind and to constitute a distinct group. With the advent of more objective methods of observation and the development of testing techniques, the presence of lesser deviates who bridge the gap between the average man and the person of rare gifts has been demonstrated. The popular concept of genius as a separate "species" probably arose in the same fashion as the similar belief regarding the feebleminded, and it is slowly being dispelled by the same methods.

The relationship between genius and eminence is a curious one. Many writers identify the two by the simple expedient of defining genius as the possession of "what it takes" to become eminent in our society. The eminent man is then considered a genius *ipso facto*. There would thus be as many kinds of genius as there are ways of succeeding in the particular society. The successful financier, for example, may be awarded an honorary university degree for his "financial genius," the victorious general for his "military genius." Society often creates a new form of "genius" in order to rationalize its allotment of eminence.

Almost any theory regarding the nature of genius could, of course, be defended by restricting the term "genius" in some arbitrary way. The broadest and most objective definition of genius is that of *an individual who excels markedly the average performance in any field*. Social evaluation, however, invariably enters into the concept. Genius is defined in terms of specific social criteria and a cultural frame of values. In our society the more abstract and linguistic abilities are

considered the "higher" mental processes. Similarly, certain lines of achievement enable the individual to earn the appellation of genius much more readily than others. Thus academic and scientific work, literature, music, and the visual arts are rated higher than, let us say, roller skating or cooking.

To be sure, very exceptional accomplishments in the latter fields might be recognized as genius, after a fashion. An internationally famed roller-skate acrobat or a renowned *chef-de-cuisine* might be called a genius and ranked higher than a mediocre scientist or painter. But in the former instances, *the attainments must be proportionately far greater* than in the latter in order that the individual may be designated a genius. And even when the term "genius" is applied to such cases, one feels that it is done only by courtesy and that the word is implicitly enclosed in quotation marks. It is apparent, therefore, that in order to have practical meaning any definition of genius must recognize the selection of significant talents which has been made within a given cultural group.

A further question which has been vigorously debated is that of *general versus specific genius*. Is the man of genius one who manifests a well-rounded intellectual superiority or one who possesses a highly specialized gift? It follows from what we know about the organization of abilities that this distinction is not a valid one. Since the intercorrelations of diverse abilities are neither highly positive nor highly negative, we should expect all degrees of generality of genius. A few individuals may excel highly in a large number of traits and thus appear to be all-around geniuses, as in the classic example of Leonardo da Vinci. Some will excel in only a few traits, and still others may have a single talent which is sufficiently pronounced to put them in the category of genius.

THEORIES ON THE NATURE OF GENIUS

Theories on the nature and causes of genius are legion. The genius has been credited with a wide variety of attributes, ranging from divine inspiration and a superhuman "spark" to imbecility and insanity. Among these diverse theories it is possible to discern four underlying viewpoints. These will be designated the *pathological, psychoanalytic, qualitative-superiority,* and *quantitative-superiority* theories.

Pathological Theories. Pathological theories [1] have linked genius with insanity, "racial degeneracy," and even feeblemindedness. Such theories date back to ancient Greece and Rome. Aristotle noted how often eminent men displayed morbid mental symptoms, and Plato distinguished two kinds of delirium: one being ordinary insanity, and the other the "spiritual exhalation" which produces poets, inventors, and prophets. The *furor poeticus* and *amabilis insania* of the Romans had reference to the same phenomenon. Democritus was among those who argued for such a relationship. It was Seneca who inspired Dryden to write his well-known line regarding great wit and madness being near allied. Lamartine spoke of the "*maladie mentale qu'on appelle génie*," and Pascal maintained that "*l'extrême esprit est voisin de l'extrême folie*." In 1836 Lélut shocked the literary world by declaring that physiological evidence furnished by the life of Socrates left no doubt but that the "father of philosophy" was subject to trances, attacks of catalepsy, and to false perceptions and hallucinations, constituting what Lélut termed "sensorial or perceptual madness." Ten years later, Lélut reached a similar conclusion about Pascal, calling attention to the latter's religious visions and hallucinations. This early work of Lélut provided an important stimulus for later theories of genius and insanity, as well as for a host of other similar analyses of the pathological traits of eminent men.

The latter half of the nineteenth century was the golden age of pathological theories of genius and witnessed the publication of many weighty tomes on the subject. Some of the leading exponents of the period were Winslow, Moreau de Tours, Mobius, Nisbet, and Nordau. This viewpoint reached its culmination in the work of the Italian anthropologist Lombroso (37, 38). His book entitled *The Man of Genius* was translated into several languages and read widely at the turn of the present century. Lombroso attributed to the genius certain *physical stigmata*, allegedly indicative of atavistic and degenerative tendencies. Among such stigmata he mentioned short stature, rickets, excessive pallor, emaciation, stammering, left-handedness, delayed development, and originality! He also maintained that there were certain similarities between the creative act of genius and the typical epileptic seizure.

Among modern exponents of modified versions of the pathological

[1] For a survey of this extensive literature, with special reference to literary and artistic genius, cf. Anastasi and Foley, 1, pp. 65 ff.

theory of genius, the most outstanding are probably Kretschmer (33) and Lange-Eichbaum (34). The former has maintained that for true genius exceptional ability is not enough. He writes, "If we take the psychopathic factor, the ferment of demonic unrest and psychic tension away from the constitution of genius, nothing but an ordinary gifted man would remain" (33, p. 28). In addition, Kretschmer applies his constitutional typology (cf. Ch. 12) to the problem of genius, arguing for a qualitative distinction between the achievements of leptosome and pyknic geniuses. The schizothyme leptosome, he claims, will tend toward subjectivity, as in lyric poetry or expressionist art; the cyclothyme pyknic, on the other hand, allegedly inclines more toward realistic painting, narrative epic poems, and the like.

The most extensive modern contribution to the pathological theory of genius has undoubtedly been made by Lange-Eichbaum (34, 35). In his *Genie, Irrsinn und Ruhm*, published in 1928, he brings together the biographies of 200 men and women of genius from all countries, periods, and fields of endeavor. All these biographies contain references to alleged abnormalities of their subjects. The reports are fully documented with a bibliography of over 1600 references, but vary in length from several pages to the simplest comment such as "for a long time psychotic." Lange-Eichbaum grants that there is not an invariable or necessary association of genius with insanity. At the same time he insists that those geniuses who have not suffered from mental abnormalities are few. Among this small minority he cites Titian, Raphael, Andrea del Sarto, Rubens, Leibnitz, and a few others. From his survey he concludes that although the proportion of the general population who are psychotic is about 0.5%, among geniuses 12% to 13% have been psychotic at least once during their lifetime. Confining his analysis to the 78 "greatest names" in his list, he finds that more than 10% have been psychotic once in their lifetime. More than 83% have been either psychotic or markedly psychopathic, more than 10% slightly psychopathic, and about 6.5% healthy. When only the 35 names representing "the greatest geniuses of all" were selected, 40% fell into the psychotic category. Over 90% were characterized as either psychopathic or psychotic, and about 8.5% normal.

Lange-Eichbaum's explanation of the association of insanity and genius is threefold. First, the pathological condition is said to increase

the strength of the individual's emotions and his responsiveness to minute stimuli, and to decrease his self-control—all of which may result in experiences which "normal" persons do not have. Secondly, Lange-Eichbaum maintains that those suffering from these conditions are likely to experience more unhappiness and feelings of inferiority, which motivate them more strongly. Finally, the tendency to a richer fantasy- and dream-life, associated with some of these disorders, may be conducive to creativity of expression.

In evaluating the evidence cited in support of pathological views of genius, several factors must be taken into account. First, in most of the studies, the evidence consists of *selected cases*. Some individuals could, of course, be found to illustrate almost any theory. The real test of the hypothesis must be based on a completely unselected sampling of geniuses. The survey of Lange-Eichbaum is probably less subject to such selective factors than many other such studies, but it is not entirely free from them.

A second point is that many geniuses may *become* maladjusted in a society built up around the average man and his needs. This is particularly noticeable in the case of a very superior child placed in a class of mediocre school children. It is probably true of superior adults too. In such a case, the maladjustment would be an *indirect result* rather than a cause or an essential component of genius. A different although related consideration is that the genius, by virtue of his superior abilities, may be more *keenly aware* of shortcomings and injustices which he observes and thus subjected to more emotional "wear and tear." It has been said that a sensitive and imaginative person cannot live as calmly as a storekeeper (61).

Geniuses, moreover, are often *regarded* as pathological by their fellow-men until the practical benefits of their work become tangible. Their undertakings are often misunderstood or ridiculed until their success is demonstrated. The familiar example of Fulton and his steamboat is a case in point. In the past, the genius has at times met with organized and violent opposition or even persecution. Life under such conditions is not very conducive to the development of a stable and well-adjusted personality. It should also be noted that, even when the genius is recognized and acclaimed as such, he is likely to be surrounded by such a glare of publicity that all his *actions and idiosyncrasies become common knowledge*. As a result, any behavioral deviation too slight to attract attention in a less outstanding individual

is pounced upon, discussed, and elaborated until it may assume the proportions of a neurotic or psychotic symptom. Finally, the *cultural setting* in which the particular man of genius lived must be considered. It is misleading to evaluate the behavior of a thirteenth- or sixteenth-century genius in terms of present criteria of abnormality. Trances and visions, for example, were not so unusual at one time in our history as they are today, nor did they have the same significance.

Psychoanalytic Theories. In common with the more recent modifications of pathological theories, psychoanalytic conceptions of genius emphasize motivational rather than intellectual characteristics (15, 22). Although admitting that a high level of ability is essential, some psychoanalysts regard this aspect of genius as a "psychological riddle" (19) and concentrate upon motivational factors. Others have taken the more extreme position that the genius does not differ in ability from the ordinary man, but differs only in what he does with his ability under strong motivational urges (63). Among the psychoanalytic concepts which have been most frequently applied to an explanation of genius are sublimation, compensation, and "unconscious processes" in creative production.

By sublimation is meant that the artistic or scientific achievement serves as a substitute outlet for thwarted drives, often of a sexual nature. The familiar illustration of the poet who composes a love lyric when he is frustrated in love comes to mind. But many of the specific cases to which some psychoanalysts have tried to apply this mechanism are much more far-fetched and seem rather forced. Compensation for real or imagined inferiorities has likewise been proposed as the principal clue to the accomplishments of genius (63). A favorite illustration is that of great orators who, like Demosthenes, developed their talent as a compensation for an initial habit of stammering or a similar speech defect. It has also been suggested that Beethoven composed his greatest works after he became hard of hearing, and that he probably had a hearing defect even in early life. As a result, his interests were allegedly centered upon auditory experiences from an early age and he began a regimen of intensive training which culminated in his outstanding musical achievements (63, p. 119). Like sublimation, compensation can probably help us to understand the motivation of some geniuses, but it should not be applied indiscriminately to all cases.

A number of creative workers, especially artists, have provided

accounts of their own creative experiences. Some of these accounts refer to production under trance-like states and to the automatic, apparently uncontrolled appearance of creative ideas. This the psychoanalysts have regarded as evidence for their theory of the importance of "unconscious processes" and the part which such processes play in creative work. The number of persons who have written such introspective accounts is, of course, small in comparison with the total number who have achieved eminence in art, science, and other fields of endeavor. Artists, by the very nature of their profession, are more likely to dramatize their own experiences than are other types of creative workers. A sobering contrast to such dramatized accounts is provided by the results of Rossman's inquiry among 710 active and successful American inventors (49). This inquiry, which was supplemented with information obtained from research directors and patent attorneys, covered both the characteristics of inventors and the nature of the inventive process. No part of this study lent any support to the popular notion of invention as a spectacular event. For this group of inventors, the creative experience was on the whole a very methodical, systematic, and matter-of-fact process.

Even among artists, those who have spontaneously written accounts of their own creative experiences may be a rather atypical group. It is likely that the more unstable, pathological individuals have, on the whole, been more interested in recording such observations, just because their experiences were more unusual and newsworthy. The records are far from factual or objective, and any preconceived theories which the individual himself may have had could have colored the original account. Finally, it should be noted that many of the psychoanalytic interpretations of the creative process as well as of the nature of genius are vague, confused, and mentalistic, often mixing literal and figurative concepts indiscriminately.

Theories of Qualitative Superiority. According to the doctrine of qualitative superiority, the man of genius is a distinct type differing from the rest of the species in the kind of ability he possesses. Such views can be distinguished from the pathological and the psychoanalytic in that they regard the man of genius as essentially *superior* to the norm. No inferiorities of any sort are implicit in this concept. The achievements of genius, according to these theories, result from some process or condition which is entirely absent in the

ordinary man. Such current expressions as "the spark of genius" reflect the popular influence of this point of view.

This approach, like the pathological, has a long history (cf. 23). In the ancient world, genius was frequently attributed to divine inspiration. The Greeks spoke of a man's "dæmon" which was supposed to possess divine powers and to furnish the inspiration for his creative work. Among those who discussed genius in these terms are Plato and Socrates. During the Middle Ages, genius was often regarded as the inspiration of a chosen mortal by the deity or by a devil, the attribution depending upon the use to which the creative talents were put.

Qualitative distinctions are also common in more recent literary and philosophical writings on the subject of genius. Mystic insights and unconscious intuitions have been attributed to the man of genius. In this connection may be mentioned the views of Schopenhauer, Carlyle, and Emerson. In psychological discussions of genius, this point of view is much less common. An example is the theory proposed by Hirsch (23), in which he differentiated three "dimensions" of intelligence. According to this theory, the first dimension is perceptual and cognitive and is shared by man and the lower animals; the second is conceptual and is common to all of mankind; the third he designates "creative intelligence" and attributes only to genius.

Qualitative distinctions appeal to the imagination of the public. The genius whom the layman acclaims differs so greatly from the rest of mankind in his achievements that he seems to belong to another species. A careful analysis of the individual's abilities, however, will reveal no essentially new process. And only a brief unbiased search discloses the presence of intermediate degrees of capacity in all lines.

Theories of Quantitative Superiority. The view that genius involves a quantitative superiority regards the genius as the upper extreme of a continuous distribution of ability. The "special gifts" and "creative powers" of genius are attributed, to a lesser degree, to all individuals. Genius is defined in terms of concrete, measurable behavior rather than in terms of unknown entities. To be sure, the accomplishments of genius are not attributed to any single talent, but to an auspicious combination of various intellectual, motivational, and environmental factors.

It follows from this view that the origin of genius is to be understood in the same terms as that of all individual differences. Many investigators, such as Galton (20, 21), Terman (55), and L. S. Hollingworth (24, 27), have placed the major emphasis upon hereditary factors. The observation that genius tends to run in families has probably given the greatest impetus to such a hereditary interpretation. The powerful environmental influences exercised by family contacts and traditions cannot, however, be overlooked. In the sections which follow, we shall examine specific findings on genius for whatever bearing they may have upon the various theories.

METHODS FOR THE STUDY OF GENIUS

Psychological investigations on the nature and development of genius have followed two fundamental approaches, viz., the study of adults who have achieved eminence and the study of gifted children. The specific procedures may be further subdivided into: (1) biographical analysis, (2) case study, (3) statistical survey, (4) historiometry, (5) intelligence test survey, and (6) longitudinal study. Although any one investigator may, and frequently does, combine more than one specific method, we shall consider them independently for clarity of presentation.

In *biographical* studies, all available published material on a given individual is examined in the effort to arrive at an understanding of the nature and origin of his genius. The investigation is limited to a single individual, who is usually chosen from the great men of the past. This method has been employed extensively by psychoanalysts, as well as by the exponents of pathological views of genius. The literature on this method runs to several thousand references (cf. 1, 34, 35).

The *case study* method consists of direct testing and observation of a single living individual. Because of the difficulty of subjecting adult geniuses to such an investigation, this method has been applied almost exclusively to gifted children. Several such studies on contemporary "child prodigies," including a number on juvenile authors, have been conducted by psychologists.

The *statistical survey* method, like the biographical, is based upon an analysis of printed records, although differing from the latter method in several essential respects. The purpose of statistical sur-

veys of genius is to discover general trends in a large group, rather than to make an exhaustive analysis of a single case. All available information on a large number of men is obtained from biographical directories, encyclopedias, *Who's Who*, and similar sources. This material is occasionally supplemented from biographies. But the former sources are employed predominantly because of the more objective, reliable, and standardized nature of their data. It will be noted that in this method the criterion of genius is chiefly eminence.

The *historiometry* method makes use of all historical material on an individual or a group of individuals. The data are gathered from a variety of sources, including biographies, directories, and original documents such as letters and diaries. The attempt is made to obtain as complete information as possible, especially on the childhood accomplishments of the great man. This material is then evaluated in terms of a more or less constant standard in order to arrive at an estimate of the individual's traits. This method was employed by Woods (71) in his study of mental and moral heredity in royalty. Terman (51) subsequently suggested an adaptation of historiometry whereby the recorded achievements are evaluated in terms of mental test norms for each age and an IQ is computed. By this method, for example, Terman estimated that the IQ of Francis Galton in childhood was approximately 200.

The *intelligence test survey* involves the direct study of large groups of intellectually superior children by means of mental tests. Extensive use is now being made of this method. The subjects are originally selected on the basis of intelligence test performance, and subsequent analyses are made with the aid of standardized intellectual, educational, and personality measures. A relatively recent development is the *longitudinal study*, in which a group of children, originally selected because of high IQ, are followed up into adolescence and adulthood.

Each of these procedures has its own peculiar advantages and disadvantages. No one can be regarded as best or poorest on all counts. The statistical, historiometry and intelligence test methods can be applied to large groups, and hence disclose general trends. They are also relatively free from selective bias, yielding fairly representative samples. The biographical and case study methods, on the other hand, give a more complete picture of the individual and enable one to note the specific interaction of various conditions in the sub-

ject's development. The study of contemporary living geniuses makes direct observation possible and avoids the judgment errors and other inaccuracies which are inevitably present in historical material. At the same time, carefully controlled observation of living geniuses offers many practical difficulties. A further disadvantage in the study of contemporaries is the possibility that the eminence of some may be short-lived and spurious and that others who are laboring in obscurity may be recognized as geniuses by posterity.

Finally, the relative advantages of studying adult geniuses and gifted children may be considered. To investigate intellectually superior children in the effort to discover the characteristics of adult geniuses seems somewhat indirect. Only a small number of such children are likely to develop into adults who can be classified as geniuses. Children, however, are available for prolonged and controlled observation and testing which would be practically impossible with adults. A further advantage of the study of gifted children is that it makes possible a developmental approach to the problem. Such an analysis may go far toward clarifying the origin and nature of genius.

STATISTICAL SURVEYS OF EMINENT MEN

Investigations of genius through statistical surveys of printed records have been conducted in England by Galton (20, 21), Ellis (18), and Bramwell (4); in France by deCandolle (14), Jacoby (32), and Odin (45); and in America by Cattell (9, 10, 11), Brimhall (5), Clarke (12), Bowerman (3), and Visher (62). Castle (8) conducted a similar survey on eminent women of all countries, but the data of this study are extremely tentative and difficult to interpret. We shall examine briefly some of the principal findings of these various surveys.

The *socio-economic background* of eminent men has generally proved to be distinctly above average. The genius who has been nurtured in a slum is the exception rather than the rule. Thus in Visher's analysis of the occupations of the fathers of 849 "starred" American men of science,[2] nearly half were found to be engaged in

[2] The "starred" men represent the most eminent persons listed in the *Directory of American Men of Science*. Those to be starred in each field of science are chosen on the basis of nominations by scientists who had previously been starred in that field. The original 1000 starred men were selected in 1903, and 250 additions were made in each new edition of the directory, prepared every five years.

the professions. This proportion is far in excess of that in the general population, the latter falling between 3% and 6%. The entire occupational distribution of the fathers of the starred men is given in Table 31.

TABLE 31 *Occupational Distribution of Fathers of 849 Starred American Men of Science*

(Adapted from Visher, 62, p. 533)

Occupational Group	*Per Cent*
Professions	45.5
Business and mercantile	23
Farming	22
Skilled labor	8
Unskilled labor	1

A similar occupational distribution is to be found among the fathers of the eminent men and women surveyed by Ellis (18). In Castle's study of eminent women of all times and nationalities, it was reported that 33.1% had fathers in the "learned professions" (8). The distribution of paternal occupation found by Cox (13) in a group of 282 eminent men and women of all countries is shown in Table 32. In this group, which covered a much earlier period in history (1450-1850), the predominance of high socio-economic level is even more conspicuous.

TABLE 32 *Occupational Distribution of Fathers of 282 Eminent Men and Women of All Countries*

(From Cox, 13, p. 37)

Occupational Group	*Per Cent*
1. Professional and nobility	52.5
2. Semi-professional, higher business, and gentry	28.7
3. Skilled workmen and lower business	13.1
4. Semi-skilled	3.9
5. Unskilled	1.1
No record	0.7

The number of *eminent relatives* may also be considered in this connection. It will be recalled (Ch. 10) that in Galton's study (20) the 977 eminent men investigated had a total of 739 known relatives who had also achieved eminence. Moreover, the closer the degree of relationship, in general, the more numerous were the eminent relatives. A follow-up of Galton's study, covering three subsequent generations and reported in 1948 by Bramwell (4), closely corroborated Galton's findings on the frequency of eminent relatives. Similar results were obtained in Brimhall's investigation (5) of family resemblance among American men of science.

TABLE 33 *Proportion of American Men of Science Born in Eastern and Midwestern States*

(From Cattell and Cattell, 11, p. 1265)

Place of Birth	*Number of Cases (per 1000 entries)*	
	1903 Group	1932 Group
Massachusetts	134	72
Connecticut	40	16
New York	183	128
Pennsylvania	66	48
Illinois	42	88
Minnesota	4	32
Missouri	14	40
Nebraska	2	20
Kansas	7	32

Certain interesting trends are suggested by Cattell's analysis of the *place of birth* of American men of science (cf. 9, 11). In his 1906 report, Cattell pointed out that cities contributed a much greater proportion of men of science than did rural sections. Although at that time the urban population was about one-sixth of the rural population, it produced a quarter of the scientific men. Even more striking is the comparison of different states which varied widely in their educational facilities. In Table 33 are shown the relative number of scientists born in each of nine states. These states were chosen as the clearest examples of a definite trend which had been operating over an interval of three decades. Corresponding figures are

shown for the original group of 1000 scientists selected in the year 1903 and for the group of 250 elected in 1932. All figures have been expressed in terms of 1000 entries to permit direct comparison.

These data suggest several conclusions which are borne out by the complete results for all parts of the country (cf. 11). In the first place, there are marked discrepancies in the relative number of eminent scientists born in different parts of the country. Secondly, these differences in birthplace correspond closely to differences in educational opportunities in various sections of the country. Thirdly, as educational facilities change, the frequency of scientists shows a corresponding change. Since the turn of the century, for example, there has been a phenomenal development of education in the midwestern states. The relative quality of education in such states has improved, new universities have been established, the contribution of state and federal funds to higher education has mounted sharply, the number of students in institutions of higher learning has increased rapidly,[3] and a powerful tradition has been built up which fosters intellectual activity. On the basis of such findings alone, we cannot, of course, draw any inferences regarding the relative contributions of hereditary and environmental factors. Whether there has been a selective migration of intellectually superior families from New England to the midwestern states, or whether the improved educational facilities have been conducive to the development of more scientists—or whether *both* of these influences have been operating—cannot be conclusively determined from the available data.

Of interest in connection with the pathological theories of genius is the relative frequency of *insanity* among the relatives of eminent men, as well as among the subjects themselves. In all statistical surveys in which the cases were not selected to prove a point, the incidence of intellectual and emotional disorders has been found to be consistently smaller among eminent men and their families than in the general population. In the group investigated by Ellis (18, p. 192), less than 2% were reported to have had either insane parents or insane offspring. Among the eminent individuals themselves, Ellis mentions 44 cases of emotional disorder out of a total group of 1030. Of these, only 13 could be definitely classed as insane during the

[3] Cf., e.g., Eells' analysis of the "center of population" of higher education from 1790 to 1920, which showed a westward movement at the rate of 60 miles per decade (17).

active period of their lives; 19 were either insane for a short period or manifested very mild disorders; and 12 developed senile dementia in old age (cf. 18, pp. 189–190).

Other facts which have been brought to light by these surveys relate to *age of parents at the time of birth of the child, order of birth,* and similar "vital statistics." It has been suggested, for example, that intellectually superior children are more often born of older parents (48). From a somewhat different angle, Lombroso (38) claimed that geniuses are the offspring of aged parents and offered this as further evidence of the pathological nature of genius. The data on this question are difficult to interpret because of the complicating factor of social level. People in the higher social classes, from which geniuses are most frequently recruited, tend to marry later and therefore have children at a later age. They also tend to have fewer children, who thus benefit all the more from educational and other socio-economic advantages. For all these reasons, parental ages are in themselves inconclusive. Among American men of science, Cattell (10, III) found 35 years to be the average age of the father at the time of the subject's birth. For English men of science, Galton (21) found the corresponding figure to be 36 years. Ellis (18) gives 37.1 years for his group of British men and women of eminence. In all these groups, however, the range of parental ages at the time of the subject's birth is extremely wide. In the majority of cases the parents were in the prime of life, contrary to Lombroso's contention.

Somewhat more conclusive is the analysis of order of birth within the family. In general the eminent individual is most often the oldest or first-born child in the family. Next in order of frequency comes the youngest child, intermediate children having the least chance of becoming eminent (cf. 18, 72). These findings are in direct contradiction to the proposed theory that older parents have intellectually more gifted offspring. It would seem that, within the same family, the superior child is most likely to be born when the parents are younger. This finding may have an environmental explanation. The first-born has traditionally enjoyed privileges in our society that his younger siblings may not have had. More is usually expected of the oldest son. If a choice must be made for economic reasons, the oldest child is usually allowed to complete his education, in preference to the younger children. These conditions might be sufficient to produce a slight degree of relationship between birth order and achievement.

Motivational factors in sibling relationships may also play a contributing part, as may the fact that the first-born probably receives more adult attention. The latter is particularly true of only children, who would all be classified as "first-born."

HISTORIOMETRY IN THE ANALYSIS OF EMINENCE

The childhood of great men, viewed retrospectively, has been the source of much controversial discussion. There is a popular belief [4] that many geniuses were dull in childhood, a number of favorite examples being cited in support of this contention. Darwin was considered by his teachers to be below average in intellect. Newton was at the bottom of his class. Heine was an academic failure, revolting against the traditional formalism of the schools of his time. Pasteur, Hume, von Humboldt, and other equally famous men were unsuccessful in their school work.

An examination of the available biographical material in such cases shows that the intellectual defect was erroneously inferred from the level of scholastic performance within a rather narrowly restricted area. The intellectually superior child may be just as maladjusted in school as the dull or borderline case. Schools adapted to the average child may be unsuited to the highly gifted pupil in many ways. The monotonous drill and rote memorization which constituted such a large part of school work in the days when men like Darwin or Hume attended school would prove particularly irksome to a bright child. Darwin, for instance, seems to have been more interested in his collections of insects than in memorizing Latin declensions, much to the annoyance of his teachers. Thus it is often impossible to accept the recorded opinions of parents or teachers regarding the intellectual status of great men in childhood.

More accurate information can be obtained from factual records of the *specific behavior* of the individual at various ages. An early attempt to conduct such an analysis of the boyhood of great men was made by Yoder (72). Fifty cases, representing a wide variety of occupations or fields of eminence were selected from the great men of six countries. All the subjects were born in the eighteenth or nineteenth centuries, except Newton, Swift, and Voltaire, who were born in the seventeenth. In general, Yoder found that ill health in

[4] Also proposed by Lombroso (38).

childhood was often exaggerated by the earlier biographers and that this condition was not so prevalent as is supposed. Feeble or delicate health may, however, offer advantages in some cases by stimulating reading and intellectual pursuits. Dickens was a good example of this. In regard to intellectual status, Yoder reports that excellent memory and vivid imagination were often exhibited by great men from early childhood.

A very detailed and comprehensive study of the childhood of great men was conducted by Cox (13), as one part of the *Genetic Studies of Genius* under the general direction of Terman. The technique employed was Terman's adaptation of the historiometry method. Through the examination of several thousand biographical references, information was gathered on the traits of 301 eminent men and women born between 1450 and 1850. Particular attention was given to childhood behavior, such as age of learning to read, letters and original compositions which may have been preserved, and early interests. Any special circumstances which might have influenced the subject's development were also noted. The material so collected was analyzed and evaluated independently by three trained psychologists. Each investigator estimated the lowest IQ compatible with the given facts for every subject, and the average of these three independent judgments was taken as the final minimum IQ estimate for the given individual.

After allowing for certain inaccuracies in the data, Cox concludes that the average IQ for the group "is not below 155 and probably at least as high as 165" (13, p. 217). The estimated minimum IQ's ranged approximately from 100 to 200. The same geniuses cited by Lombroso and others as instances of early mental inferiority were invariably found to give evidence of high IQ's during childhood. Among these may be mentioned Lord Byron, Sir Walter Scott, and Charles Darwin, whose estimated childhood IQ's proved to be 150, 150, and 135, respectively. Among those receiving IQ's above 180 were Goethe, John Stuart Mill, Macaulay, Pascal, Leibnitz, and Grotius.

Another interesting finding pertains to the average estimated IQ of persons achieving eminence in different fields (53). Philosophers topped the list with a mean IQ of 170; next came poets, dramatists, novelists, and statesmen with 160; scientists had a mean of 155, musicians 145, artists 140, and military leaders 125. This hierarchy probably reflects at least in part the close association of "intelligence"

with verbal aptitude in our present standards of evaluation. Those groups with mean IQ's of 160 or over were engaged in activities in which written or spoken language played a predominent role. Farthest from the verbal field in their area of accomplishment are the persons at the bottom of the list: military leaders, artists, and musicians.

In the same survey, 100 geniuses were selected for whom the relevant records were especially full, and ratings were assigned to each person on a number of specific intellectual, emotional, and character traits. These ratings, like the IQ's, were based upon the childhood behavior of the subjects, and the averages of two independent raters were used. As a group, the subjects proved to be unquestionably superior in all the traits rated, and were especially outstanding in such characteristics as desire to excel in their efforts, steadfastness of effort, persistence in the face of obstacles, intellectual work devoted to special pursuits, profoundness of apprehension, and originality and creativeness. Another sub-group of 50 cases, similarly selected because of fullness of data, were rated in a like manner for physical and mental health in childhood. The distribution of the group in these respects is reported to be fairly normal and to show no greater per cent of unfavorable deviants than are found among unselected school children.

Some of the inconsistencies and confusions regarding the association of "genius" and "insanity" may result from the common use of these blanket categories as though they represented single entities (41, 64). If we ask *what kind* of genius and *what kind* of abnormality, we are more likely to get a significant and consistent answer. Re-analyses of the original Cox data, for example, have shown that the incidence of emotional abnormalities is greatest among the "aesthetic type" (poets, novelists, artists, musicians) and the "reformer type" (revolutionary statesmen or radical religious leaders). It is least among scientists, soldiers, statesmen, and conservative religious leaders. The more "imaginative" genius is likely to show more psychopathic characteristics than the eminent "man of action." As for specific types of abnormality, analyses of the same group suggest that introversion, emotional excitability, and fanatical self-confidence are the most frequent. Considering how often these geniuses were right in their novel ideas, the last-mentioned symptom seems to be more indicative of fanaticism in the rest of mankind than in the genius!

Other investigators have corroborated these findings regarding the *specificity* of the "genius personality." One survey (47) compared 120 men of science with 123 men of letters, both groups having lived during the nineteenth century. The literary group was limited to poets, novelists, and dramatists; the scientists included only workers in the biological and physical sciences and in mathematics. One interesting difference was found in the socio-economic backgrounds of the two groups. Though both the scientists and the literary men came chiefly from the professional class, the two groups differed in that the scientists were much more likely than the men of letters to come from the farmer and artisan class. For the men of letters, the socio-economic class which ranked second in frequency to the professional was the semi-professional. On the other hand, actual poverty was more often reported for literary than for scientific men. The scientists as a group were described as more cheerful, modest, and sociable. The literary men excelled in persistence, but were also more emotional, gave more evidence of neurosis, and had a slightly poorer health record both in childhood and adulthood. Also relevant is a recent survey (50) of the characteristics of "research workers," conducted by a similar method. Biographical material was examined for 250 research workers ranging from Euclid and Pythagoras to contemporary living scientists. Among the characteristics found most frequently were creativeness, enthusiasm, and aggressiveness; least frequent were religiousness, self-control, and good health.[5]

The results of all these studies have to be accepted with caution because of possible weaknesses in the procedures. Much depends upon the representativeness of the samples, the fullness of the available data, and the objectivity and accuracy with which the recorded behavior items are evaluated by the investigator. When great men of the past are considered, a certain amount of historical perspective is also required, in order to judge the individual against his own cultural setting. On the whole, however, such studies do show that the men and women who achieved eminence tended to come from favorable environments, gave early indication of superior ability, and were not as a group appreciably more unstable than the less gifted. At the same time, it should be clear that "the genius" is not one but many kinds of person.

[5] Cf. also Rossman's survey of the dominant characteristics of inventors, as reported by the inventors themselves (49).

THE GIFTED CHILD

The "Child Prodigy." Since geniuses have generally displayed superior talents in childhood, a direct study of gifted children should prove fruitful in an analysis of genius. The traditional or popular concept of the "child prodigy" is that of a weak, sickly, unsocial, and narrowly specialized individual. His achievements are expected to be of the nature of intellectual "stunts" and to have little or no practical value.

One of the earliest recorded cases of such a child prodigy is that of Christian Heinrich Heineken, whose achievements are described by his teacher in an old German book published in 1779 (cf. 29, 46). At the age of 10 months this child was able to name objects in pictures; before 12 months he had memorized many stories in the book of Moses; and at 14 months he knew the stories of the Old and New Testaments. At 4 years of age he could read in his native language, had memorized 1500 sayings in Latin, and also knew French. At this time he was able to perform the four fundamental arithmetic operations, and he knew the most important facts of geography. His fame spread throughout Europe and he was summoned to appear before the King of Denmark. True to the traditional picture, however, Christian Heinrich was a sickly child, and at the age of 4 years-4 months he died.

Contrary to popular belief, the case of Christian Heinrich is not at all typical. As an example of a highly gifted child who developed into a healthy and successful adult we may consider the case of Karl Witte (cf. 66). Born in Lochau, Prussia, in 1800, this "child prodigy" lived until he was 83, having retained his excellent intellectual powers to the end. Karl was literally educated from the cradle. His father was convinced of the efficacy of early training and undertook to prove this with his son. The child was never taught "baby talk." All the games he played were games of knowledge. When only 8 years old, he read with apparent pleasure the original texts of Homer, Plutarch, Virgil, Cicero, Fénelon, Florian, Metastasio, and Schiller. He matriculated as a regular student at Leipzig at the age of 9. Before his fourteenth birthday he was granted a Ph.D. degree. Two years later he was made a Doctor of Laws, being at the same time appointed to the teaching staff of the University of Berlin.

Karl Witte's father, in discussing the boy's education, wrote:

> . . . he was first of all to be a strong, active, and happy young man, and in this, as everybody knows, I have succeeded. . . . It would have

been in the highest degree unpleasant for me to have made of him pre-eminently a Latin or a Greek scholar or a mathematician. For this reason, I immediately interfered whenever I thought that this or that language or science attracted his attention at too early a time (66, pp. 63–64).

Karl seems not to have been in the least vain or spoiled. He never paraded his knowledge, was modest and unpretentious, and not infrequently tried to learn from his companions what they knew better than he. He had many playmates of his own age and we are told that "he got along so well with them that they invariably became very fond of him and nearly always parted from him with tears in their eyes" (66, p. 187).

Contemporary case studies of gifted children by psychologists likewise lend no support to the view that such children are necessarily inferior in other respects. In 1942, L. S. Hollingworth brought together in one book (29) 31 case reports of children whose IQ's were over 180. Such IQ's should occur about once in over a million cases. The accomplishments and adjustment of children in these IQ levels are illustrated by the following cases.

A gifted juvenile author, Elizabeth ——, obtained a Stanford-Binet IQ of 188 when tested at the age of 7 years-10 months (cf. 54; 55; 29, pp. 35–37). She ranked high in all other intellectual and educational tests, but showed a special interest and talent for the composition of prose and poetry. This child was reported to be in excellent health and free from physical defects; she was a year or so accelerated in physical development. Elizabeth's superior linguistic abilities were apparent from an early age. At 19 months she could express herself clearly and also knew the alphabet. By her eighth birthday she had read approximately 700 books, including such authors as Burns, Shakespeare, Longfellow, Wordsworth, Scott, and Poe. By this age she had also written over 100 poems and 75 stories. The following is a specimen of her literary products, written at the age of 7 years-11 months and entitled "Fairy Definition":

> Fairies are the fancies of an imaginative brain
> Which wearying of earthly realities aspires to
> Create beings living only in thought
> Endowing the spirits thus created
> With all genius for giving Happiness.

A case which attracted wide attention in the 1920's is that of a boy known in the psychological literature as E—— (29, pp. 134–158).

When first tested at the age of 8 years-11months, E—— obtained a mental age of 15–7, which gave him an IQ of 187. He also did well on all other tests except those involving manual dexterity. He is reported as being strong and healthy, but not much inclined to indulge in games and sports. At the age of 12 he was admitted to Columbia College. On the Thorndike Intelligence Examination for High School Graduates, he ranked second among 483 competitors. During his freshman year at college all his academic grades were B or better, with the exception of physical education, in which his grade was C. He is described as being a "good sport" and getting along well with the other students. He received his A.B. degree at 15, being also elected to Phi Beta Kappa. At 16 he obtained his M.A. degree, and by 18 had completed practically all requirements toward the Ph.D. degree except the dissertation. On the CAVD Intelligence Examination, his score was 441, which falls approximately in the upper ¼ of 1% of college graduates. Thus, during the period over which he was investigated, E—— showed no tendency to drop below the high intellectual level indicated by his initial IQ.

These cases are typical examples of intellectually superior children. Exceptional talents in childhood are not incompatible with good health, physical vigor, longevity, or a well-rounded personality. To be sure, puny, timid, and sickly children can be found among the gifted, as among the intellectually normal or dull. But such cases are very few and cannot be regarded as representative of the group as a whole.

The highly gifted may, of course, have their own special adjustment problems, especially during childhood and adolescence, by virtue of their exceptional intellectual status. But such maladjustments are an indirect *result* of high intellect, rather than a cause or an intrinsic component of genius. Among the possible problems encountered by the child whose IQ is much above 150 (28, 29) are those arising from the fact that he is younger and hence *smaller and weaker than his classmates*. This condition may make him more susceptible to bullying and may interfere with his participation in athletics and active games. A second source of difficulty is the "*isolation*" from contemporaries and from the common activities of others which is likely to result when the individual's interests and abilities are so unlike those of his fellows. *Negativism* toward authority may develop when the child realizes that authority is often irrational or erroneous

in its operation. *Intolerance* and unwillingness to "suffer fools gladly" may follow observations of relatively inept thinking on the part of associates. The superior child may also develop habits of *inefficient work and laziness* because ordinary school work offers no challenge to him. Such work habits may carry over into later educational and even vocational activities.

For these reasons, L. S. Hollingworth (28, 29) concludes that the optimum IQ from the viewpoint of personal adjustment, leadership, and acceptance by one's fellows—with the "accompanying emoluments and privileges" which such acceptance entails—falls between 130 and 150. To be sure, the adjustment difficulties of the highly gifted child are of the sort that can be prevented by proper understanding and a suitable educational environment (52). During the past two or three decades the special education of the gifted child has made rapid strides,[6] a progress to which L. S. Hollingworth herself made some of the most outstanding contributions (25, 26, 29). The outlook for even the most highly gifted "prodigy" need not, therefore, be a pessimistic one.

Intelligence Test Surveys. The testing of large groups of intellectually superior children has revealed the *continuity* which exists between the average child and the highly gifted "prodigy." In order to include a sufficiently large number of cases in such studies, the standard of selection must be lowered. But by surveying a wider range of superior intellect a more complete picture will be obtained. Since the rise of the mental testing movement, a number of studies on moderately large groups of superior children have appeared (cf. 40). The most extensive project of this sort is that begun in 1921 by Terman and his associates, and reported in the *Genetic Studies of Genius* (cf. 55, 6, 58). Because of the more comprehensive nature of this study and its essential agreement with the findings of other investigations, it will be described in greater detail.

The total group employed in Terman's study (cf. 58, Ch. I) included 1528 California children, ranging in IQ from 135 to 200 and in age from 3 to 19. These children represent approximately the upper 1% of the school population. Of these, 661 elementary school

[6] As early as 1924, The National Society for the Study of Education devoted one of its *Yearbooks* entirely to teaching methods suitable for gifted children. For a survey of more recent developments in this field, cf 40, 43, 60, 69, 70. Attention is also called to the recently formed American Association for Gifted Children, which is specially concerned with the problems of the gifted child (65).

children constituted the "main experimental group," on which the major findings of the initial test survey were based. This group was compared, in an extensive series of tests and measures, with control groups composed of random samplings of school children. For reasons of expediency, different control groups were employed for various comparisons, the number of cases in such groups ranging from about 600 to 800.

The *socio-economic level* of the gifted group was decidedly superior. Among the fathers of the gifted children, 31.4% belonged to the professional class, 50% to the semi-professional or higher business class, 11.8% to the skilled labor class, and 6.8% to the semi-skilled or unskilled labor class. The average school grade reached by the parents of the gifted group was 11.8, and by the grandparents 10.0. In comparison to the average person of their generation in the United States, the parents in this group had received from 4 to 5 grades more schooling. Moreover, a third of the fathers and 15.5% of the mothers had graduated from college. The number of eminent relatives and ancestors was also far in excess of that which would be expected by chance, and many of the families had highly distinguished genealogies.

The homes of the gifted children were visited by field workers, and were rated from 0 to 6 on necessities, neatness, size, parental conditions, and parental supervision.[7] The average rating of the homes was over 4.5 in each of these five categories, and only 10% of the homes received a total rating which was distinctly poor. Neighborhood ratings and income level were also considerably better than the generality for California.

We may next consider certain *vital statistics* as well as medical and physical data obtained on the gifted children themselves. The frequency of insanity in the family was lower than average. Only 0.4% of the parents and 0.3% of the grandparents and great-grandparents had a record of insanity. As in the studies on adult genius, the gifted group contained a greater proportion of first-born children than the general population. The gifted children developed at a more rapid rate than the normal from early infancy. They walked on the average one month earlier and talked 3½ months earlier than the control groups. The onset of puberty was also somewhat earlier than normal. Physicians' examinations showed superior health and relative

[7] The Whittier Scale for Grading Home Conditions was used for this purpose.

freedom from defects in the group as a whole. Similarly, such conditions as "nervousness," stuttering, headaches, general weakness, and poor nutrition were less common in the gifted than in the control groups. In height and weight, physical and muscular development, and strength, the overlapping of gifted and control groups was almost complete. Such differences as did occur, however, favored the gifted group.

The *educational accomplishments* of the gifted group were, of course, far in advance of the normal.[8] About 85% of the gifted children were accelerated and none was retarded. The administration of standardized achievement tests in school subjects revealed that the majority of these children had already mastered the subject matter from one to three grades above that in which they were located. Thus with reference to his actual *abilities*, the gifted child is often retarded rather than accelerated in school-grade location. The gifted children as a group tended to excel in all school subjects; one-sidedness was not characteristic of these children. Their superiority was greatest, however, in such subjects as language usage, reading, and other "abstract" work, and least in shop training, sewing, cooking, and similar "craft" subjects.

The gifted group displayed a wide range of *interests* outside of their school work, as well as an active *play life*. A two-month reading record kept by the children showed that the gifted read more than the control at all ages. At 9, the number of books read by the gifted group was three times that of the control. The range of topics covered was also wider and the quality of the books superior in the gifted group. Similarly, the gifted children were more enthusiastic, had more intense interests in general, and reported more hobbies than the control group. Collections were nearly twice as common among the gifted as among the control, and tended also to be larger and more often of a scientific nature. A questionnaire on play information showed that the typical gifted child of 10 knew more about playing and games than the average child of 13. Apart from the fact that the play interests of the gifted children were more mature than those of the control children of their own age, no conspicuous differences were found in their play activities.

In *character* and *personality development*, the gifted children were

[8] This was partly the result of the method of selection. Teachers were asked to name the brightest children as well as the youngest child in each class, and from among these the gifted subjects were chosen by intelligence tests.

also found to be in advance of the normal. This was confirmed both by scores on objective tests of emotional adjustment and character traits and by parents' and teachers' ratings. On a specially devised battery of seven objective personality tests, the differences in favor of the gifted group were large and significant in every test.[9] From about 60% to 80% of the gifted group equaled or excelled the average of the control group in each of these tests.

The findings of the California study have been closely corroborated by studies on similar groups in the Middle West by Witty (67, 68), in New York by L. S. Hollingworth (24, 27, 29), and in England by Duff (16). Superior home and parental background, better-than-average health and physique, outstanding educational achievement, and greater emotional maturity and stability were characteristic of all these gifted groups.

THE GIFTED CHILD GROWS UP

Among the many superstitions entertained in regard to geniuses and child prodigies is that which claims that the gifted child deteriorates as maturity is approached and that his ultimate mental level will be average or even inferior. Prolonged case studies on a few individuals, as well as a number of scattered investigations on groups of gifted children, have quite conclusively disproved this view.

The most extensive follow-up of a large group of gifted children is that conducted under the direction of Terman, and reported in Volumes III and IV of the *Genetic Studies of Genius* (6, 58). An integral part of the plan of the California study included periodic follow-ups of the original group of gifted children. The first follow-up, after six years, involved the retesting of small samples of the original subjects, as well as a detailed progress report on a larger proportion of the group. At this time, most of the subjects were in their adolescent years. The high school records of the group were fully as distinguished as their performance in elementary school. Achievement tests, as well as intelligence tests, showed continued superiority, as did also general health and personal and social adjustment. Participation in extracurricular activities and leadership among classmates were especially outstanding for the group.

[9] Critical ratios (diff./σ_{diff}) ranged from 3.87 to 14.41. Critical ratios of 3 or more indicate that the chances of a true difference are over 99/100.

Subsequent follow-ups were conducted in 1936, 1940, and 1945. The 1936 follow-up [10] was a preliminary questionnaire survey, whose findings were superseded and rendered obsolete by the two later follow-ups. The 1940 follow-up was a thorough and comprehensive one, involving an extensive testing and interviewing program by field investigators. At this time, the average age of the group was 30 years. Of the original total of 1528 children, 61 were deceased in 1940 and 33 could not be traced. The remaining 1434 cases participated in the intensive survey. In 1945, a supplementary follow-up was conducted by mail. By this time the majority of the group were 35 years old, an age at which adult careers are clearly taking shape. The results of the 1940 and 1945 follow-ups, taken together, constitute the basis for the analysis of the adult status of the gifted group, reported by Terman and Oden in *The Gifted Child Grows Up* [11] (58). Plans are under way for the further continuation of this extensive longitudinal study.

Adult intellectual status was measured by a specially constructed *Concept Mastery Test* consisting of opposites and analogies and covering many fields of information. Through this test it was possible to estimate that the average adult IQ of the gifted group was about 134, representing a drop of 17 points from their childhood average of 151. The authors show that such a drop is no greater than would be expected from regression (cf. Ch. 8). Such regression, however, would result not only from errors of measurement in the Stanford-Binet and the Concept Mastery Test, but also from differences in the functions measured by the two tests, as well as from actual behavior changes in the subjects resulting from maturation or learning. In other words, predictions over a twenty-five-year period are subject to considerable error, not only because of the unreliability of the tests, but also because much can happen to change the subjects during such a period. The important point, however, is that the obtained change in this group is not significantly different from that expected by chance and gives no evidence of any special decline of ability.

Educationally, the gifted group excelled in all comparisons. They attended college in much larger numbers, took graduate degrees much more often, and received better grades and many more academic

[10] This is the follow-up whose results were reported by Terman and Oden in 1940 (56, 57).

[11] This book is Volume IV of *Genetic Studies of Genius*.

honors than any other groups with which comparisons were made. Among the men, 69.5% completed college, and among the women 66.8%. The per cent receiving Ph.D. degrees was over five times as large for the men and over eight times as large for the women in the gifted group as in a representative sampling of college graduates. A special study of educational acceleration in the gifted group not only showed acceleration to have been common, but also lent no support to the view that such acceleration may be detrimental. Any slight social handicap suffered by the very accelerated subjects during adolescence seems to have been fully overcome in later years. In fact, whatever differences were found in later achievement or adjustment tended to favor the accelerated group.

TABLE 34 *Occupational Classification of Gifted Men and of All Employed Men in California (1940)*

(From Terman and Oden, 58, p. 172)

	Occupational Group	*Per Cent of Gifted Men* (N = 724)	*Per Cent of Employed Males in California (1940)* (N = 1,878,559)
I.	Professional	45.4	5.7
II.	Semi-professional and higher business	25.7	8.1
III.	Clerical, skilled trades, retail business	20.7	24.3
IV.	Farming and other agricultural pursuits	1.2	12.4
V.	Semi-skilled trades, minor clerical	6.2	31.6
VI.	Slightly skilled trades	0.7	17.8
VII.	Day laborers: urban and rural	0.0	

In occupational level, the gifted group stood far above the average, being represented in the higher professions by eight times its proportional share. In Table 34 will be found the occupational distribution of the gifted men, together with the corresponding distribution of all employed males in the 1940 California census. Nearly half of the gifted men are in the professional category, as contrasted to less than 6% of the generality; the corresponding proportions in the semi-professional and higher business category are 25.7% and 8.1%, respectively. On the other hand, only 6.2% of the gifted men are in semi-skilled trades, as against 31.6% of the generality. Similarly, less than one per cent of the gifted group are in the slightly skilled trades and none in the unskilled, as contrasted to 17.8% of the generality

in these two classes combined. Even in comparison with groups of male college graduates, the gifted group excels markedly in occupational status.

The occupational history of the gifted women is much more difficult to interpret, since jobs and careers have a very different significance for the two sexes in our culture. Of the entire group of gifted women, 42% were housewives and not gainfully employed. Only 48% reported full-time employment at the time of the follow-up. Among those employed, the largest number (30.8%) were in secretarial or other office work, and the second largest (21.1%) in elementary or high school teaching. Social work, arts, writing, and college teaching and research each claimed from 5% to 7%. Perhaps the most outstanding finding in the comparison of the gifted women with other groups of women college graduates is the smaller proportion of the gifted who chose teaching and the larger proportion who chose office work. The interpretation of any of these results would be hazardous, in view of the multiplicity of factors which influence the occupational activities of women in our society today. The discussion of sex differences in the next two chapters may help to clarify some of these results.

The mortality rate of the gifted group was below that of the generality, and both physical and mental health remained superior. The incidence of delinquency, alcoholism, and serious maladjustments was less than in the general population, and there was considerable evidence of good emotional and social development and breadth of interests. Participation in extracurricular activities was as conspicuous in college as in high school. Hobbies and avocational interests were well developed and closely resembled those of any contemporary American group. An active interest in political and social matters is suggested by the fact that 91% of this group reported that they voted in all national elections, in contrast to only about 70% in the general California population. The social and political attitudes of the gifted group showed no marked deviation from the generality. The subjects' war records, in both military and civilian capacities, were also found to be quite creditable and distinguished.

Of considerable interest are the data on marital status and marital adjustment. The incidence of marriage among both the gifted men and the gifted women is above that of college graduates of the same age, and is about equal to that in the general population. Intelligence

tests of the spouses as well as the offspring showed both to be quite superior, but below the average of the gifted group itself. On specially designed tests of "marital aptitude" and "marital happiness," the present group was somewhat superior to other groups less highly selected in intelligence. Sexual adjustment was in all respects as normal as in less gifted groups. Divorce rate was no higher than in the generality of comparable age.

A special study of individuals whose initial IQ's had been 170 or higher showed them to compare favorably with the rest of the gifted group. They were more often accelerated in school, received better grades, and continued their education longer than the average of the entire group. They were as well adjusted emotionally and more successful vocationally than the rest of the group. Thus it seems that this particular group of exceptionally gifted children were, on the whole, able to overcome the special problems and difficulties which their high intellectual level might engender.

Probably one of the most interesting analyses in the entire survey is the comparison of the 150 men rated "most successful" (Group A) with the 150 rated "least successful" (Group C) in adult achievement. Despite the high average accomplishments of the entire gifted group, the adult achievement of individual members ranged from "international eminence to unskilled labor" (58, p. 311). In the effort to clarify some of the correlates of adult achievement, the two contrasted groups A and C were compared on about 200 items of information which had been secured between 1921 and 1941. The most conspicuous differenc were the superior educational and vocational level of the parents of the "A" men, as well as the greater "drive to achievement" on the part of the "A" men. For example, over 50% of this group had fathers who were college graduates, in contrast to 15.5% of the "C" men. More than twice as many fathers of the A's were in the professions. As for the subjects themselves, both self-ratings and ratings by family and associates showod the largest A-C differences in "integration toward goals," "perseverance," and "self-confidence." Significant differences in favor of the "A" men were found in school acceleration, the A group graduating from elementary, high school, and college at younger ages. Initial IQ's also averaged significantly higher for the A group; but this difference was not large, the two averages being 155 and 150. In summary, factors related to home background seemed to play a major role in the

adult achievement of these men, all of whom were within the upper levels of intelligence. Among such men, motivational factors—themselves probably traceable to environmental conditions—often made the difference between outstanding achievement and mediocrity.

From an over-all view of such follow-up investigations on gifted children, what can we conclude? At the outset, it should be noted that some corroboration of the California findings, although on a much smaller scale, is provided by follow-ups of the New York (30, 31, 39) and midwestern (68) groups cited previously. These studies, too, indicate that the gifted child, on the whole, grows up to be an intellectually superior and fairly well-adjusted adult.

There is a possibility that the California results may be unduly optimistic about the emotional adjustment of children in the highest IQ levels. Perhaps the method of *selecting the original group* may be partly responsible for such a finding. The major group was chosen on the basis of teachers' recommendations, the children thus recommended being then given intelligence tests for the final selection. As a check on this procedure, the entire population of three schools was tested, following the teacher nominations. The results showed that about 90% of all the children who qualified for the study on the basis of test scores would have been reached by the usual procedure. It is possible that the 10% who were thus lost to the study may have included a disproportionately large number of scholastically and emotionally maladjusted cases. Their exclusion from the study by the method of search might thus lead to an unduly optimistic picture in these two respects.[12] The possible effects of *participation in the study* upon the subjects' subsequent development should also be considered. No control subjects were followed up along with the gifted subjects. Not only the knowledge that one is a "gifted child," but the personal interest in each subject which was apparently shown by the field investigators and project directors cannot be completely discounted. The experimental design employed in the study includes no control for this factor.

In reference to the question of what constitutes "genius" and how

[12] Corroborative evidence is provided by a recent test survey of over 45,000 children in grades 4 to 8 (36). Without knowledge of test results, teachers were asked to indicate each child whom they considered "a distinct problem," "extremely mentally retarded," and "a genius." The influence of the child's classroom adjustment and academic interests and achievement was evident in the choices for the "genius" category.

the eminent adult is related to the gifted child, several points may be noted. First, it has been pointed out by L. S. Hollingworth (27) that an IQ of 135 or 140 is certainly far below the "genius level." Such IQ's fall within the upper quarter of college students. In fact, in some of the better colleges, the mean IQ is close to 150. On the basis of her own follow-ups of groups of gifted children, Hollingworth proposed that an *IQ of 180 or higher* is more nearly at the "genius level," equipping the individual for academic and professional distinction, original and creative work, the winning of prizes, and other evidences of eminence.

Any definition of genius, however, is so intimately related to the specific cultural setting that to consider the individual apart from the time and place in which he lived and worked is highly artificial. It is quite generally agreed, moreover, that a high IQ alone is no guarantee of "genius." Some writers (2, 7) have particularly emphasized the role of special aptitudes, such as talent in art, music, or mechanics, in their definition of genius. The importance of motivational and emotional factors, stamina, environmental background, and opportunity has also been repeatedly stressed. Many have been impressed with the quality that so often makes the genius undertake and persist in what others have labeled "impossible." It was this point that Bolitho eloquently expressed when he wrote: "Where common sense is horrified, where the sign 'impossible' is raised in warning, kindness or spiteful joy, there is your exit, prisoner; there is the door of adventure." [13]

CHAPTER

18

Sex Differences: Basic Problems

SPECIALIZATION OF VOCATIONAL ACTIVITIES with regard to the sexes has been a powerful social tradition in almost all cultural groups. The particular tasks assigned to each sex vary from group to group and are even occasionally reversed, but some differentiation of activity is practically universal. These distinctions are impressed upon the individual from early childhood, either by actual overt differences in training and play activities, or by the more subtle but perhaps more effective inculcation of traditional beliefs and ideals. It is apparent that in most societies the effectual environments of the two sexes are fundamentally diverse from an early age. Under such conditions, we should expect pronounced variation in the emotional and intellectual development of the two sexes. By a curious circular argument, however, these socially conditioned behavioral differences are often attributed directly to innate factors.

The belief in hereditary sex differences in intellectual and emotional traits is an old and persistent one. It is only since the development of objective and quantitative testing methods that the notion of "female inferiority" has been dispelled among scientists. In the general public, this belief still prevails, as is manifested by the reluctance to open certain educational and professional opportunities to women and by the frequent discrimination against individuals on the basis of sex alone. The reasoning underlying such practices is that it would be futile to provide identical training for men and women, since the existing differences in their behavior are so clearly apparent. This view, of course, fails to consider the possibility that the existing sex differences may themselves be the result of the diverse training and environment of the two sexes.

The objective study of sex differences in intellectual or personality

traits began shortly after the rise of the mental testing movement. In 1910 Woolley, one of the first investigators in this field, listed less than a dozen psychological studies on sex differences (52). A review appearing in 1926 contained over 200 such titles (26). In 1935, one bibliography included more than 300 studies (33). Today, the relevant investigations number well over a thousand and the entire field is rarely covered in a single survey. Every type of function has been surveyed, from sensori-motor processes, through simple perceptual and associative tasks, to more complex intellectual activities and personality characteristics. Almost all tests, shortly after their construction, have been administered to men and women, and their scores compared. It was a relatively easy task to gather such data, especially after the advent of standardized group tests; but it was quite a different matter to determine what these data meant.

In common with other group comparisons, the study of sex differences in behavior presents a number of methodological difficulties. An understanding of these problems is essential for the proper interpretation of the findings of any reported study. For this reason, we shall begin by considering the basic questions which must be raised in the evaluation of any data on group differences.

EVALUATION OF GROUP DIFFERENCES

Selective Factors. In all group comparisons, selective factors may operate to vitiate the results. When a group is not a random or representative sample of the population from which it is drawn, it is said to be a select group. Such a sampling is unsuited for any type of investigation, since any results obtained with it could not be generalized but would apply only to the specific group employed. An additional complication in the comparison of two populations arises from the fact that selection may have operated differently in the two groups. Thus if a group of college girls were compared with trade school boys, the two samplings would be selected in different ways. Not only is neither group representative of men or women in general, but the one represents the upper end of the female distribution and the other a central or slightly inferior segment of the male distribution with respect to education and correlated variables. In addition to being unrepresentative, these groups are not comparable.

Selective factors are often difficult to detect and usually difficult to

control. An example of such a selective factor is provided by the comparison of high school boys and girls. Offhand, we might say that groups of boys and girls attending the same high school constitute truly comparable samples for the study of sex differences. But investigations on elementary and high school students have demonstrated that this is not the case.

Let us examine, for example, two independent studies in which the Pressey Group Test of Intelligence was administered to 2544 elementary school children between the ages of 8 and 16 (37) and to 5929 high school seniors ranging in age from 16 to 23 (5). The percentages of boys who reached or exceeded the *median score* of the girls, as well as the number of cases in each group, are shown in Table 35. In the elementary school study, the data are reported separately for each age group. In the study on high school seniors, a single summary

TABLE 35 *Sex Differences in Intelligence Test Scores of Elementary School and High School Groups*

(Adapted from Pressey, 37, p. 327, and Book and Meadows, 5, p. 61)

Elementary School Group Age	*Number of Cases* Boys	Girls	*Per Cent of Boys Reaching or Exceeding Girls' Median*
8	57	92	40
9	132	154	34
10	176	177	42
11	179	167	41
12	182	180	44
13	174	174	39
14	138	162	43
15	102	139	41
16	62	97	49
High School Seniors	2422	3503	56.2

figure is given for the entire group. It will be noted that in the elementary school grades the girls excel at all ages, although the sex difference is negligible among the 16-year-olds. Among the high school seniors, however, this relationship is reversed, over 50% of the boys reaching or exceeding the girls' median score.

This reversal becomes intelligible if we examine the relative number of each sex in the elementary grades and in the senior year of high

school. Throughout the high school period there is a much more rapid elimination of boys than girls. Boys whose academic work is not satisfactory are more likely to leave school and go to work, whereas girls tend to be kept in school longer. Girls also seem to adjust better to the school curriculum and school routine in general. The less intelligent girls will exert more effort and manage to pass sufficient subjects to stay in school, while boys in the same situation are more likely to rebel against school work. This explanation was borne out by an examination of the academic history of those students who had dropped out of high school. Owing to the differential action of this selective influence upon the two sexes, differences between the intelligence test scores of high school boys and girls cannot be regarded as true sex differences. In the evaluation of any study on group differences, selective factors are one of the most subtle forms of error to which we must be constantly alerted.

Significance of a Difference. One of the first questions which the psychologist asks regarding any reported group difference concerns the statistical significance of the difference. The problem of significance arises from the fact that in any investigation only a *sample* of the entire *population* is employed. For example, if the population under investigation is defined as public school children in American cities, data may be gathered on some 5000 or 6000 children in a dozen schools. From these results, the investigator generalizes to the entire population. If the sampling was carefully chosen to be representative of the given population, such conclusions will not be far in error. The figures thus obtained, however will not be identical to those which would have been secured by testing the entire population of American city public school children. Nor will the results from successive samplings of the population coincide perfectly. Had another sampling of 5000 city public school children been employed, slightly different results would have been obtained.

This variation in results from sampling to sampling within the same population is known as a *sampling error*. Statistical measures of *reliability* provide a theoretical estimate of the probable limits within which such errors will fall. Formulas are available for the computation of the sampling error of all statistical measures, such as averages, differences between averages, measures of variability, and correlation coefficients. It is thus possible to estimate the maximum amount of

variation to be expected in any statistical measure if the experiment were repeated on another sampling of the same population.

When we ask, "Does this mechanical aptitude test show a significant difference in favor of the boys?" we mean simply this: "Would the boys' average score still be higher than that of the girls if we were to test the entire population of boys and girls from which our samples were drawn?" We refer to the difference we actually found as the "obtained" difference, and to the difference in the entire population in which we are interested as the "true" difference.

By simple formulas,[1] we can compute the standard error of the difference ($\sigma_{diff.}$) and with it, the *critical ratio* or "t." The latter is simply the ratio of the obtained difference to its standard error ($t = diff./\sigma_{diff.}$). It has been customary for many years to regard a critical ratio of 3 or higher as evidence that the obtained difference is significant. In other words, when the obtained difference in favor of, for example, the boys is 3 or more times as large as its standard error, we can be virtually certain that there is a "true" difference in favor of the boys in the entire population.

More recently, there has been a tendency for statistical workers to express the significance of a difference more precisely in terms of the actual *probability* of a true difference in favor of one or the other group. With a t of 3, the probability that the obtained difference indicates a true difference is about 99.7 out of 100. For the probability to be exactly 99/100, the critical ratio would have to be 2.58. In such a case, the chances that the population difference favors the same group which excelled in our tested sampling are 99 out of 100, and the probability that the difference is either absent or in favor of the other group is only 1 out of 100. This is the basis for the frequently encountered statement that the difference is "significant at the .01 level of confidence." Another way of expressing the same conclusion is to state that the probability that the obtained difference resulted from chance factors alone and does not indicate a true group difference is .01 or less ($P < .01$).[2]

A hypothetical example will serve to illustrate the use of such tests of significance. Let us suppose that a group of sixth grade school

[1] Cf. any recent textbook on psychological statistics, such as Garrett, 13, Ch. VII.

[2] With very small samples, the critical ratio must be considerably larger than 2.58 to permit the same level of confidence in the results. For further treatment of these technical details, cf. Garrett (13).

boys and girls obtain the following average scores on an intelligence test:

Girls	85
Boys	80
Difference	5

Let us suppose further that we have computed the σ_{diff} and found it to be 4. The critical ratio of the obtained difference will then be 5/4 or 1.25. Since this is less than 2.58, the study has not demonstrated the presence of a true sex difference at the .01 level of confidence. The 5 points in favor of our group of girls may result from chance factors, and another investigator, giving the same intelligence test to other samples of boys and girls, may find a difference in favor of the boys or perhaps no difference at all. We can, in fact, estimate [3] that the probability that our obtained difference resulted from chance factors is 21/100. This probability is considerably higher than the customary 1/100 at the .01 level of confidence.

We may consider one further illustration, with the following data:

Boys' average	130
Girls' average	110
Difference	20
σ_{diff}	4

In this example, $t = 20/4 = 5$. Since this is much larger than 2.58, we know that the likelihood that the obtained difference has arisen from chance factors is much less than 1/100. The sex difference is therefore clearly significant at the .01 level of confidence ($P < .01$). We are safe in concluding that there is a true difference in favor of the boys. The chances of our being wrong in such a conclusion are less than 1 out of 100.

The standard error of an obtained difference depends upon the *size* of the samplings employed as well as upon the amount of *variability* within the samplings. It is apparent that the larger the sampling, the more reliably will the results be established. If the sampling were infinitely large, the standard error would be zero, since the entire possible population would then have been included. In most of the earlier investigations on sex differences, the samples employed were so small as to yield extremely large standard errors, had the latter been computed. The sex differences reported in such studies may thus have been due entirely to chance errors.

[3] By reference to a table of the normal probability curve.

Similarly, the wide variability existing within each sex in regard to any given trait renders the differences between averages less reliable. If all women were of identical height, for example, and all men were likewise equal in height, then sex differences in height could be reliably established by comparing only one representative of each sex. All other samplings would yield the same difference, since variation within each sex would be zero. The greater the variability within either group, the larger will be the standard error of the obtained values. In the computation of SE's, both the number of cases and the variability of the group are taken into account.[4]

Overlapping. When Samuel Johnson was asked which is more intelligent, man or woman, he replied, "Which man, which woman?" This is a vivid way of expressing the wide individual differences within each sex, with the consequent overlapping between their distributions. Since in any psychological trait women differ widely from each other, and men also vary widely among themselves, any relationship found between group averages will not necessarily hold for individual cases. Even when one group excels another by a large and significant amount, individuals can be found in the "inferior" group who will surpass certain individuals in the "superior" group. Owing to the large extent of individual differences within any one group as contrasted to the relatively small difference between group averages, an individual's membership in a given group furnishes little or no information about his status in most traits.

In most discussions of group differences, attention has been focused primarily upon averages. For a complete picture of the relative standing of the two groups, however, some index of the degree of overlapping should be included. The best procedure would be to report the entire frequency distributions of the two groups. This is often impracticable, however. A simpler alternative, in the case of normally distributed samplings, is to state the percentage of subjects in one group who reach or exceed the median (or average) of the other. Complete overlapping would then be indicated if 50% of one group reached or exceeded the median of the other.[5] If more than 50% of

[4] The standard error of a mean is found by the formula: $\sigma_M = \frac{SD}{\sqrt{N-1}}$. The standard error of the difference between two means is in turn based upon the standard errors of the two separate means.

[5] The curves will not coincide, of course, if the *ranges* are unequal. In such a case, complete overlapping is obtained only in the sense that one distribution is contained entirely within the other. Moreover, if either distribution is pronouncedly asymmetrical, such a measure of overlap may be misleading.

group A reach or exceed the median of group B, then group A is to that extent superior to group B; if less than 50%, A is inferior to B. Occasionally, some other value is substituted for the median as the point of reference. Thus the investigator might report the percentage of group A which reaches or exceeds the highest score obtained in group B, or the percentage of group A which reaches or exceeds the upper quarter of group B.

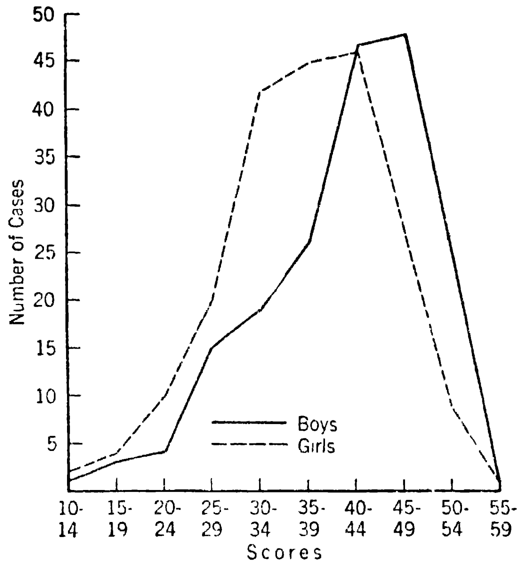


Fig. 87. Distribution of Boys and Girls on a Test of Arithmetic Reasoning. (Data from Schiller, 42, p. 67.)

That the establishment of a statistically significant difference between two groups does not preclude the possibility of extensive overlapping between them is illustrated in Figure 87. This figure gives the distribution curves of 189 boys and 206 girls in the third and fourth elementary school grades on a test of arithmetic reasoning. The average score of the boys is 40.39 and that of the girls 35.81. The difference between the averages is 4.58 points and the standard error of this difference is only 0.85. The difference is thus over five times as large as its standard error and can be regarded as significant with a high degree of confidence. An examination of the distribution curves, however, reveals extensive overlapping between the two groups, a

very large percentage of boys and girls falling within the same range of scores. Moreover, 38% of the girls obtained scores higher than the boys' average, and 24% of the boys scored below the girls' average.

Nature of the Measuring Instrument. It is a platitude to insist that, in order to obtain significant data on any question, an accurate measuring instrument must be employed. Yet the methods of measurement employed in the study of sex differences, as well as in other group comparisons, have frequently been crude and often wholly unsuited to the problem. Thus *ratings* by associates were used in many of the earlier investigations on sex differences and especially in those concerned with personality traits. Teachers' ratings of school children were especially common. It is obvious that such ratings do little more than reflect the systematic bias of the judges. In the comparison of such groups as the sexes or various "races" or nationalities—about which popular stereotypes exist within each culture—ratings cannot be regarded as an index of the subject's actual standing.

The reliability of the tests (cf. Ch. 2) should also be taken into account. If a test is too short or if performance on it is affected by too many irrelevant factors, it will yield different results on repeated administrations. On such a test, the scores of the *same individuals* will vary widely from time to time. These discrepancies in test scores are known as *errors of measurement.* Group differences found with a short and poorly constructed test may be entirely spurious and may be expected to disappear upon a re-examination of the same subjects.

Much confusion has also been introduced into discussions of group differences by the relatively loose designations assigned to most psychological tests. If a test is labeled "analytic reasoning," there is a tendency to assume that it actually measures that trait, although such a trait may not even exist as a unitary function and may consist of a manifold of independent abilities. Similarly, if two tests are given the same name, they are commonly regarded as measuring the same function. A hypothetical example will show how this practice may affect group comparisons. Let us suppose that one investigator has constructed a sentence completion test, which he labels a measure of "logical thinking." In such a test, as in most verbal tests, girls will probably excel. If now another investigator also sets out to construct a test of "logical thinking" and decides to employ arithmetic problems as his material, he will find that boys excel in this trait. The results of

the two studies will thus seem to be in direct contradiction, owing to the use of a common term to cover two discrete types of behavior.

Many discrepancies in the data on sex differences may be attributed simply to such a confusion of terminology. Unless identical tests are administered in an identical manner, we cannot assume that the same functions were measured in every case. The use of a different time limit, for example, might change a power test into a speed test and thus yield entirely different results. A slight alteration in the directions might make it more difficult for the subjects to understand what is required of them and might thereby introduce a new element into the test, viz., ability to follow verbal instructions. "Intelligence" scales are probably the best example of the use of general terms in describing widely diverse tests. Much controversy has been occasioned by the application of such scales. Owing to the employment of "intelligence" scales which sample different sets of abilities, some students of sex differences have concluded that boys were more intelligent, others that girls were more intelligent.

A closely related problem pertains to the use of "lump scores" in group comparisons. Group differences in specific abilities may be completely obscured by the comparison of total or average scores on a battery of tests. If, for example, boys excel in numerical aptitude and girls in verbal aptitude, and a scale of so-called general intelligence is weighted equally with items from both fields, no significant sex difference in total score will be found. Should the scale be overweighted with items of one type, on the other hand, it will favor the group excelling in that trait, and will indicate an apparent difference in general intelligence. In recent years, with the development of factor analysis, there has been a growing tendency to look for group differences in separate abilities rather than in "general level of performance." In the study of group differences, it is of the greatest importance to state results in specific terms and to limit conclusions to the particular materials, procedure, and other conditions of each investigation.

SEX DIFFERENCES IN ACHIEVEMENT

The relative intellectual achievements of men and women through the ages have frequently been cited as evidence of a sex difference in ability. An examination of any biographical directory or encyclopedia

shows a far greater number of men than women to have achieved eminence. And of the few women listed in such compendiums, many acquired fame through special circumstances, such as royal birth, rather than through the possession of exceptional talent. In Ellis' study (12) of British genius, only 55 women were included in the total group of 1030 subjects. Nor did the standard of eminence seem to be higher for women than for men. On the contrary, Ellis claims that many of the women in his group had become famous "on the strength of achievements which would not have allowed a man to play a similarly large part" (12, p. 10). Cattell's carefully prepared compilation of the 1000 most eminent persons in the world listed only 32 women. Of these, 11 were hereditary sovereigns and 8 became eminent through misfortune, beauty, or some other circumstance. This leaves an extremely small number who may be said to have distinguished themselves through their superior talents (8, p. 375).

Similar results were obtained by Castle (7) in her statistical study of eminent women. A total of 868 names of women were collected, representing 42 nations and covering a wide range of epochs from the seventh century B.C. to the nineteenth century. The *largest number* of women in the group achieved eminence through literary pursuits, 337 or 38.8% of the subjects being classified in this field. The *highest degree of eminence*, however, as indicated by the number of lines allotted to the individual in standard biographical directories, was obtained by women as sovereigns, political leaders, mothers of eminent men, and mistresses. Among the other non-intellectual factors through which women achieved fame in the past are listed marriage, religion, birth, philanthropy, tragic fate, beauty, and "immortalized in literature."

In more recent times, the discrepancy in number of men and women who have distinguished themselves in intellectual pursuits is still large, although constantly diminishing. In the 1933 edition of *American Men of Science* (cf. 9, p. 1264), 725 women were listed out of a total of 9785 entries in the pure sciences. The percentage of women in the various fields ranged from 2.1% in physics to 22% in psychology. In the group of 250 scientists who were newly "starred" [6] in this edition, only 3 women were included (9). In fact, out of a total of 2607 scientists starred between 1903 and 1943, only 50 were women (50).

[6] Cf. footnote 2 in Chapter 17.

The interpretation of such achievement data is obviously complicated by the many factors besides ability which determine eminence. The recorded differences in achievement *could* be fully accounted for in terms of the environmental conditions which have prevailed. Many types of *occupations* have been completely closed to women until recently. Thus, on the basis of their sex alone, women have been effectively barred from achieving eminence in a number of fields. When women have eventually been admitted officially to such vocations, prejudice and discrimination against them have still been so prevalent that only a few could succeed. Even today, competition is not on an equal basis for men and women in most occupational fields.

Educational opportunities have likewise been very dissimilar for the two sexes (cf. 14), although at present the environments of the two sexes are more nearly equated in this respect than in any other. Institutions of higher learning were slow to open their doors to women. Although America was in advance of most other countries in the education of women, until nearly the middle of the nineteenth century there was not a single institution of collegiate rank in this country which admitted women. Professional and post-graduate education was not available until a much later date. Even in the elementary and secondary schools, the traditional curriculum of girls was different from that of boys, including much less science and more literature, art, and other "genteel" subjects.

Nor can *general home influences* be disregarded. Even in the most enlightened and progressive homes, differences are introduced in the environments of boys and girls which may prove very important in determining subsequent behavior development. In general, girls are considered weaker and more frail than boys; they are sheltered more and are taught to be neater and quieter than their brothers. Boys and girls are given different toys to play with and different books to read. All these apparently minor environmental factors, operating constantly and from a very early age, may exert a lasting influence upon the development of the child's interests, emotional characteristics, and intellectual talents.

Finally, the relatively intangible but highly effective factor of *social expectancy* should be mentioned. This operates to perpetuate all group differences, once they have been established. What is expected of an individual is a powerful element in the determination of what he will do. When such expectation has the force of social tradition behind it

and is corroborated at every instant by family attitudes, everyday contacts in work and play, and nearly all other encounters with one's fellow-beings, it is very difficult not to succumb to it. As a result, the individual himself usually becomes convinced that he is "superior" or "inferior," or that he possesses this or that talent, interest, or attitude, according to the dictates of his particular culture.

Perhaps the follow-up studies of gifted children, discussed in the preceding chapter, may offer a clue to adult sex differences in achievement. In the California study (47), it will be recalled, the adult occupations of the women were on the whole quite undistinguished. The number of women engaged in careers of university teaching, research work, art, or writing was quite small. The reported sex differences in adult vocational activities are especially noteworthy when we remember that the men and women in this group had been so selected as to fall within *the same IQ range in childhood.* Moreover, initial IQ showed a fairly close relationship to occupational level among the men, but virtually no relationship among the women.[7] In fact, two-thirds of the women with IQ's of 170 or above were housewives or office workers.

The statistics on higher education also favored the men in this group. Although the percentages graduating from college were closely similar for the two sexes, many more men than women took graduate degrees, especially at the doctoral level. The influence of cultural traditions, social pressure, and the common conflict between marriage and a career can be recognized in the follow-up of this group of intellectually superior women. Such a study may help us to understand some of the reasons why gifted women more rarely achieve eminence than do gifted men.

SEX DIFFERENCES IN VARIABILITY

During the last decade of the nineteenth century, the doctrine of sex differences in intellectual variability[8] rose to prominence. It was

[7] Women's scores on the Concept Mastery Test administered in the adult follow-up did show a significant relationship to occupational status, but this may have been largely a result of educational differences. The fact that the women in the upper occupational levels had necessarily continued their education longer may itself have enabled them to do better on the Concept Mastery Test (47, p. 184).

[8] The possibility of greater male variability in physical traits was originally alluded to by Darwin, although he does not seem to have considered the problem of great importance (cf. 36).

pointed out that, although the *average ability* of men and women might be equal, the distribution of ability in one sex might cover a wider range than in the other. Thus it was suggested that the variability of intelligence among males is greater than among females, there being more men than women at either extreme of the distribution. These hypothetical distributions are illustrated in Figure 88. It will be noted that, theoretically, the averages of two groups can be identical while the ranges differ considerably.

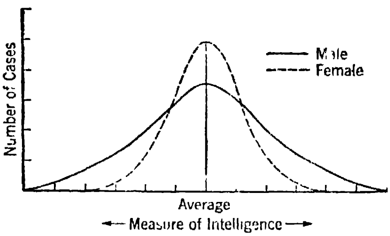


Fig. 88. Hypothetical Distribution of Intelligence among Men and Women according to the Doctrine of Greater Male Variability.

The doctrine of greater male variability was regarded as a fundamental biological law and was believed to hold for all traits, physical as well as psychological. Thus Havelock Ellis, one of its chief protagonists, wrote as follows:

> From an organic standpoint, therefore, women represent the more stable and conservative element in evolution (11, p. 421) . . . in men, as in males generally, there is an organic variational tendency to diverge from the average, in women, as in females generally, an organic tendency, notwithstanding all their facility for minor oscillations, to stability and conservatism, involving a diminished individualism and variability (11, p. 425).

This doctrine enjoyed a long popularity and was accepted by a number of psychologists during the first quarter of the present century (cf., e.g., 8, 48). The evidence offered in support of the greater intellectual variability of the male was twofold. On the one hand, the statistics on eminence were cited as proof of the greater frequency of superior intellect as well as of the presence of more extreme positive deviants in the male sex. Similar data were presented to establish the wider range of male intelligence at the lower end of the distribution. Surveys of institutions for the feebleminded in several countries revealed a consistent excess of males among the inmates. Thus it was argued that there were more idiots as well as more geniuses among men, and that women as a group tended to cluster more closely around the average or mediocre degrees of ability.

The cultural basis of sex differences in the attainment of eminence has already been discussed. No biological law need be invoked to account for the greater frequency of men in the biographical directories and encyclopedias. The greater incidence of males in institutions for mental defectives has likewise been shown to result from cultural factors. This was especially demonstrated in a study by L. S. Hollingworth (16) on 1000 cases referred for examination to a psychological clinic in New York City, as well as 1142 cases in residence at a New York City feebleminded institution. Analysis of intelligence test scores and other available data revealed the differential operation of a selective factor in the case of the two sexes.

In the first place, the males referred for examination, as well as those actually committed, were on the average much younger than the females. Secondly, the IQ's of the females presented for examination were lower than those of the males. This difference in IQ was even greater when the cases actually committed were compared. A survey of the previous occupations and general case histories of the subjects suggested that the probable explanation of these findings lies in the uncompetitive nature of many occupations open to women. This makes the detection of feeblemindedness as well as the necessity of commitment less likely among women than among men. A girl of moron level can survive outside of an institution by turning to housework, prostitution, or marriage as a means of livelihood. The boy, on the other hand, is forced into industrial work at a relatively early age and will soon reveal his mental deficiency in the severe competition which he encounters. Thus, although there is an excess of males in institutions for mental defectives, it would seem that there are more feebleminded females outside of institutions.

A similar differential selection has been found to operate in admissions to special classes for mentally retarded children in the public school system. In a survey conducted in Baltimore (4), results showed that about three times as many boys as girls were enrolled in such special classes. The remaining girls of corresponding ability, however, were found in regular public school classes.[9] Apparently the differences in social and economic conditions met by the two sexes have led to a "double standard" in the classification of boys and girls as mentally retarded.

Karl Pearson (36) was among the first to challenge the adequacy

[9] Cf. also Rigg (39)

of studying sex differences in variability by a comparison of the extremes of the distribution. He called attention to the need for direct measurement of *variability around the average* in large groups of unselected subjects. Pearson himself computed coefficients of relative variability for several classes of data, consisting chiefly of physical and anatomical measurements on adults. He found no evidence of greater male variability, but rather a slight tendency toward greater female variability. Similarly, Hollingworth and Montague (18) collected a large number of physical measurements on 1000 male and 1000 female infants *at birth,* thus ruling out any possible effects of differential environment. No consistent sex difference in variability was found.

A mass of data is now available on male and female variability in a wide variety of traits (cf. 29, 38). In such characteristics as height, weight, physiological maturity, dentition, and anatomical development, the data are inconsistent. The relative variability of the two sexes differs with the specific trait under consideration, the age of the subjects, their social and economic level, and even the particular community in which the data are obtained. Intelligence test results exhibit a similar lack of consistency. On individual tests such as the Stanford-Binet, no sex difference in variability is generally found; on many group tests, boys are slightly more variable. Age is also a factor in determining the relative variability of the sexes on intelligence tests. The same is true of variability on special aptitude tests as well as in school achievement. The findings differ with the specific situation, in one case the boys being more variable, in another the girls. In the large majority of cases furthermore, the differences in variability in favor of either sex are too slight to be of much significance.

In recent years, it has been possible to check the theory of sex differences in variability on very large and representative samplings. For example, in a Scottish survey in which all children born on any of four specified days were given the Stanford-Binet, data were obtained on 444 boys and 430 girls with a mean age of 10 years-5 months (27). The sex difference in variability in this group was negligible and insignificant, the critical ratio of the difference being less than 1. Similarly, in an American survey, 5069 boys and 5010 girls in grades 3 to 8 of 22 city schools were given the National Intelligence Test (40). No significant sex difference in variability was found with any of the measures of variability employed. An extensive investigation of this question was also conducted on American high school and col-

lege students, utilizing data which had been collected in a survey of 49 Pennsylvania colleges and a number of Pennsylvania high schools (38). Several different comparisons were made on intelligence and achievement tests as well as on certain physical measures. All groups used in these comparisons included over 1000 persons of each sex. In this study, too, neither sex was found to be consistently more variable, the results differing not only with the area of measurement, but also with the measuring instrument employed.

Another approach has been to compare the relative frequency of boys and girls at the extremes of the distribution of intelligence test scores. The California study of gifted children has sometimes been cited in support of the theory of greater male variability, since more gifted boys than girls were located in the survey. The total group included 857 boys and 671 girls. Among the children with IQ's of 170 or over, there were 47 boys and 34 girls. On the other hand, in L. S. Hollingworth's compilation of case studies of children with IQ's over 180, 16 girls and 15 boys were found (17). Witty's group of 50 Kansas City children with IQ's of 140 or higher included 24 girls and 26 boys (51).

It should be noted that the children in the California study were located in large part through teachers' recommendations. Those in Witty's group were found by administering a group test of intelligence to the entire school population in grades 3 to 7 in Kansas City, Missouri. The Hollingworth cases were identified either through their conspicuous achievements or through intelligence tests administered for other reasons. It is thus likely that the excess of boys in the California group resulted from the effect of sex stereotypes on teachers' judgments. Perhaps a girl with a high IQ was more often regarded by her teachers simply as a "good pupil," while a boy with the same IQ was judged to be "brilliant."

Such an explanation in terms of selective factors is supported by the results of complete school surveys with intelligence tests. In a study in which the National Intelligence Test was given to all the children in grades 3 to 8 in 22 city schools, the percentage of boys did not differ significantly from the percentage of girls in the combined upper and lower 7% of the entire distribution (49). There were, however, more girls in the upper 7% and more boys in the lower 7%. In a more recent survey (23) with the Kuhlmann-Anderson Intelli-

gence Test, in which approximately 45,000 children in grades 4 to 8 in 36 states were tested, the upper 10% of the group likewise included an excess of girls (2676 girls *vs.* 1853 boys). Among the highest 2% of the distribution, girls again predominated in the ratio of 146.3:100. Among the lowest 10%, the reverse tendency was found, there being 3009 boys and 1618 girls (28). The largely verbal content of most intelligence tests, as well as their dependence upon school work, probably gives the girls an advantage and accounts for their superior performance.[10] There is, however, no evidence in these surveys for a greater male variability, nor for a greater frequency of boys at the upper IQ levels.

One additional point should be considered in connection with the relative frequency of boys and girls at high IQ levels. With increasing age, gifted boys are more likely to retain their high IQ or even to show a rise, while girls with the same initial IQ's are much more likely to show a drop (10, 25, 47). In the California study, for example, the excess of gifted boys was much greater in the high school than in the elementary school sampling. Moreover, in the follow-up testing of boys and girls within the same initial IQ ranges, the girls showed a greater mean drop than the boys during adolescence as well as adulthood. Several explanations could be suggested for such a finding. The content of intelligence tests may favor girls more at the younger ages. Or girls may develop more rapidly in intellectual functions and the boys may "catch up" as they grow older. One plausible explanation is that, with increasing exposure to traditional activities and social pressures, the intellectually superior boy will on the whole continue to improve in intellectual functions, while the equally superior girl is more likely to be steered into less intellectual pursuits. Sex differences in educational, vocational, and avocational activities would in turn be reflected in an increasing divergence of the intelligence test scores of the sexes with age.

SEX DIFFERENCES IN INFRAHUMAN ANIMALS

Since cultural factors so often complicate the interpretation of observed sex differences in human behavior, it may be of interest to examine sex differences in infrahuman species. It has been argued

[10] Cf. pp. 651 ff. and 660 ff.

that if similar sex differences in behavior are observed at various phyletic levels, such differences are more likely to be directly or indirectly traceable to a structural basis. In maze performance, as well as in other *learning* tasks, sex differences in animals are inconsistent and negligible (35, 49). Although the experimental data on many of the higher forms of animals are quite meager, there seems to be no evidence for a sex difference in ability. What differences have been found pertain rather to emotional characteristics.

There is a considerable body of data—from field studies, the observations of animal breeders and trainers, and the descriptive accounts of laboratory workers—all of which indicate greater *aggressiveness* in the male of most species (6, 15, 41, 44, 53). Fighting, restlessness, and resistance to control have been commonly reported as more characteristic of male than of female animals. That this may be related to the presence of male sex hormones is suggested by a number of experiments involving the removal of gonads, as well as the injection of sex hormones. It is not only reproductive behavior which is affected by such endocrine factors, but also other behavior characteristic of one or the other sex, such as pugnacity or singing in certain species of birds (20, 34, 44, 46).

On the other hand, we must guard against overgeneralizing from such results to sex stereotypes in the human. Animal data which do not fit the familiar human stereotypes can also be found. Carefully controlled studies on *timidity* in rats, for example, showed females to be less timid than males (2). This sex difference persisted, although to a reduced degree, after the removal of gonads from rats of both sexes. In another series of investigations with rats, the female was found to be *more active* than the male (cf. 49). Also contrary to the traditional human stereotype were observations made on the *mating behavior* of a certain species of monkey (6), in which either sex may initiate the sexual advances preparatory to copulation. There is no indication that the male of this species necessarily takes the initiative in this respect.

All in all, the available findings on sex differences in animal behavior must be interpreted with considerable caution. Such observations may provide leads for the investigation of possible physiological correlates of behavioral characteristics, but it would be premature to make any generalizations regarding universal sex differences in any behavioral function.

THE ROLE OF PHYSIOLOGICAL FACTORS

General Physical Status. Some of the animal experiments have suggested the part which sex hormones may play in the general behavior of males and females, over and above the role of these hormones in reproductive behavior. The fact that endocrine secretions are carried to all parts of the body through the blood stream has led to considerable speculation regarding the broader behavioral effects of sex hormones. It should be noted that in terms of sex-hormone production, there is not a sharp contrast between males and females, but the difference is rather one of degree. All males, besides secreting the male sex hormone, androgen, also secrete some female sex hormone, estrogen. Similarly, all females secrete some androgen along with estrogen. It is the relative proportion of the two which determines the degree to which the individual develops masculine or feminine characteristics.

Another possible source of general sex differences is provided by the sex-determining chromosomes themselves. It will be recalled (Ch. 4) that every cell in the body receives a complete set of chromosomes. For the female, each body cell contains 23 pairs of chromosomes plus an XX pair; for the male, each body cell contains the same 23 pairs plus an XY pair. In this respect, then, the two sexes differ *in every cell of the body*. This does not mean, of course, that every body cell must necessarily develop differently in men and women, since not all genes may be active in the development of every cell. But these sex differences in gene constitution, repeated in every body cell, may provide a mechanism to account for many of the physical differences between the sexes.

Sex differences have, in fact, been reported for almost every physical variable, including body build, minute anatomical characteristics, physiological functioning, and biochemical composition (46). Moreover, the difference in most of these respects increases with age. Thus the human male averages approximately 5% heavier than the female at birth and 20% heavier by age 20; in height, the male excess increases from about 1% or 2% in childhood to about 10% by age 20.[11] Muscular strength shows a consistent difference in favor of males at all ages (21, 46). From early infancy, males likewise exhibit

[11] During a few years in early adolescence, girls are on the average taller and heavier than boys, but this results from the *developmental acceleration* of girls, to be discussed in the subsequent section.

greater "muscular reactivity," as illustrated by a stronger tendency toward restlessness and vigorous overt activity. Of possible relevance to such an excess of muscular reactivity is the greater mean vital capacity [12] of males. This difference is especially significant because vital capacity is an important factor in sustained energy output. In early childhood, the average vital capacity of boys is about 7% greater than that of girls; by adulthood, the male excess reaches about 35%. The vital index, or ratio between vital capacity and body weight, is likewise greater for males at all ages at which measurements have been made. Thus, even in proportion to his body weight, the human male consumes more fuel and produces more energy than the female.

All these physical differences may play an important part in sex differences in play activities, interests, and achievement in various fields of work (46). It is reasonable to expect, for example, that the greater strength and motility of boys increase the likelihood of their manipulating mechanical objects, and thus indirectly facilitate the development of clearer mechanical concepts. Aggressiveness and dominance in social relations may likewise be initially fostered by greater body size, strength, and endurance.

Rate of Maturation. It has been clearly established that girls not only reach physical maturity earlier than boys, but that throughout childhood they are farther advanced toward their adult status in physical development (41, 45, 46). Several investigators have compared the *height* and *weight* of boys and girls at successive ages. In order directly to compare the developmental status of the two sexes in these traits, each age average can be expressed as a percentage of the adult norm for that sex. In Table 36 will be found such percentages for boys and girls between the ages of 6 and 17, the figures being based upon data from several investigations. It will be noted that *at each age measured*, the girls have attained a greater percentage of their adult height and weight than the boys. Similar results were obtained in an extensive investigation by Baldwin (3), in which the *same subjects* were measured at successive ages. At certain ages the developmental acceleration of the girls is so great that they are actually taller and heavier than boys, in absolute measures. In Baldwin's data, the girls were found to be superior in height between the ages of 11 and 13, and in weight between 9 and 16.

[12] Vital capacity is the total volume of air that can be expelled from the lungs after a maximal inhalation.

TABLE 36 *Percentage of Final Growth Which Has Been Attained at Ages Preceding Maturity*

(From Lincoln, 21, p. 20)

Age	Height		Weight	
	Boys	Girls	Boys	Girls
17.5	100	100		
16.5	97.5	99.2	100	100
15.5	94.5	98.3	88.7	95.1
14.5	90.3	96.3	78.9	87.4
13.5	86.4	93.3	70.0	79.0
12.5	83.4	89.4	63.5	70.0
11.5	80.6	85.6	58.4	61.8
10.5	78.0	82.5	54.1	56.0
9.5	75.1	79.3	49.0	51.0
8.5	73.3	76.1	45.0	46.7
7.5	69.1	72.8	40.9	42.4
6.5	65.9	69.0	37.4	38.5

Other aspects of physical development show a similar acceleration of the female sex. It is a well-known fact that girls reach *puberty* earlier than boys, the difference averaging from 12 to 20 months in various groups. *Skeletal development* can be measured by the relative degree of ossification, or hardening, of the bones in different parts of the body. In this also, girls have been found to be in advance of boys at every age (cf. 41, 46). A similar difference has been found in *dentition*. In general, girls shed their deciduous teeth sooner and get their permanent teeth at an earlier age than boys. In the case of certain teeth, these differences amount to one year or over (41, 46). The general developmental acceleration of girls begins *before birth*. Girls are on the average more mature than boys at birth and there is some evidence which indicates that they tend to be born after a shorter gestation period than boys (41).

The significance of sex differences in the rate of physical growth has been emphasized by several writers (cf., e.g., 5, 24, 36, 41). It has been suggested that girls may be accelerated in intellectual as well as physical development. If this were the case, equated age groups of boys and girls would not be comparable. It would then be necessary to equate the sexes in regard to developmental stage or physical maturity rather than chronological age. But such a procedure would introduce

an inequality in amount of training and general environmental stimulation. This problem, of course, arises only in the comparison of children, and does not apply to adults. Children, however, have been the most frequent subjects for surveys on sex differences, both because of their greater accessibility in large numbers and because they have been exposed to a relatively more homogeneous environment.

It should be noted that intellectual acceleration of girls has not been directly demonstrated. Its possibility has only been inferred by analogy with physical development. It is doubtful, however, whether physical maturity can have much influence upon intellectual development. The data on the relationship between psychological and physical traits are too consistently negative for such an assumption (cf. Ch. 12). In emotional and other personality traits it is probable that the onset of puberty and the relative physiological maturity of the individual introduce an uncontrolled factor in sex comparisons at certain ages. But in regard to the individual's intellectual status, the environmental stimulation to which he has been exposed is far more significant than slight differences in physical condition.

Another possible implication of the developmental acceleration of girls is a social one (41). Because of their physical acceleration, adolescent girls tend tc associate with boys older than themselves. This probably accounts also for the usual age discrepancy in marriage. Since the girl is generally younger than the boys with whom she associates—and younger than the man she marries—she is surpassed by most of her male associates in education, intellectual development, and general experience. Such a situation may well be at the root of many social attitudes toward the two sexes. A younger individual is likely to have less wisdom, information, and sense of responsibility than an older one, and such an age difference may have been interpreted and fostered as a sex difference.

Viability and Physical Defects. At all ages, the female shows more "viability," or capacity to maintain life, than does the male. The interpretation of mortality statistics in adulthood and even in later childhood is, of course, complicated by differential hazards met by the two sexes in their traditional occupational and recreational activities. That the higher mortality rate of males cannot be explained wholly on this basis is, however, indicated by several facts.

First, prenatal and infant deaths are more common among boys

(41, 46). It has been estimated that the ratio of male to female conceptions lies between 120:100 and 150:100. Although 20% to 50% more boys are conceived,[13] however, only 5% or 6% more boys than girls are born. Thus even before birth, death has already taken a much greater toll from the male sex. At every stage of prenatal development, the percentage of male deaths is greater than that of females. Moreover, this difference in viability is not limited to the human, but

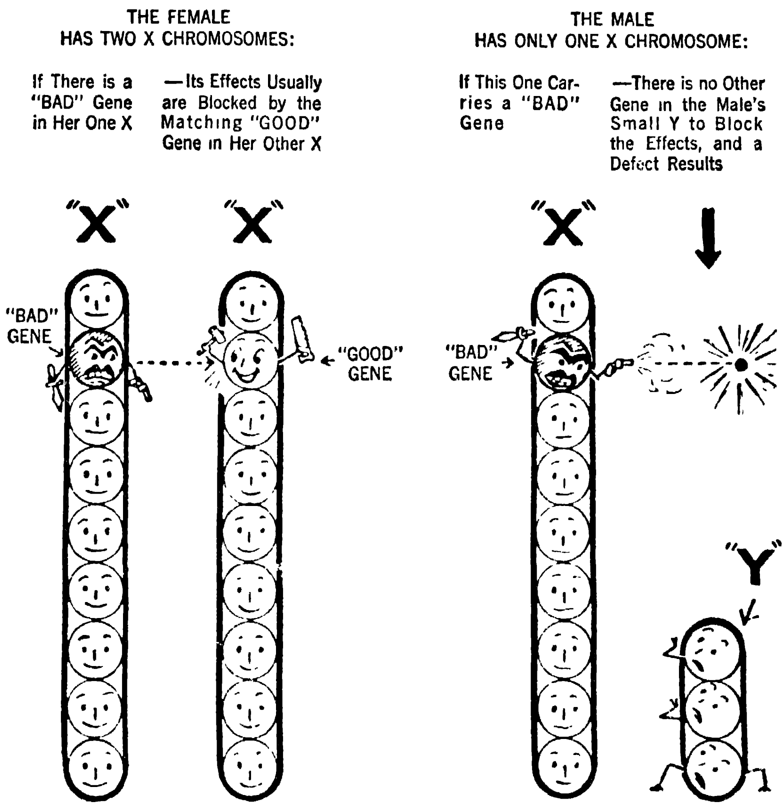


Fig. 89. Dramatized Schema Illustrating Why Males Have More Hereditary Defects. (From Scheinfeld, 41, p. 60.)

[13] The reason for the excess of male conceptions is not clear. It has been suggested that the male-producing, or Y-bearing, spermatozoon is lighter and more motile than the X-bearing spermatozoon. Another possibility is that the male-producing spermatozoon has a better chance of survival in the uterine environment for either chemical or physical reasons.

is true for lower animals as well. Throughout life, the male appears to be biologically more vulnerable in many ways. He is more susceptible to infection and is more often afflicted with physical defects. All but a very small number of defects are more common among males.

One reason for this sex difference in viability and in physical disorders may be found in the sex chromosomes. Since the female receives two X-chromosomes, the effect of a defective gene in one of these chromosomes may be counterbalanced by a normal gene in the other. This relationship is illustrated in the dramatized diagrams shown in Figure 89. The male, on the other hand, receives only one X-chromosome. The Y-chromosome contains relatively few genes, and it is doubtful whether any of them are counterparts of X-genes. It is thus much more likely that a defective gene in the male will find no normal counterpart to check its effect (Fig. 89). This relationship between corresponding genes in each pair of chromosomes can perhaps be best understood when we realize that a defective gene is probably one lacking in certain essential chemical substances. Such a deficit can be overcome by the presence of the same substance in the corresponding normal gene.

We could speculate at length regarding the possible social implications and indirect psychological effects of the greater viability of the female. For example, one result is the increasing excess of women at the upper age levels—a condition which influences the relative opportunity for marriage. A proportional scarcity of males makes marriage a more competitive undertaking for the female than for the male. This situation could in turn be reflected in divergent personality development in the two sexes.

Homeostasis. An interesting physiological concept which has received increasing attention in discussions of sex differences is that of homeostasis, or the *stability* of bodily functions. There is considerable evidence suggesting that homeostatic mechanisms, which tend to keep the body in its normal condition, operate within narrower limits in the male (46). Thus men show less fluctuation in such measures as body temperature, basal metabolism, acid-base balance of the blood, and level of blood sugar. The observation that females are more subject to flushing and fainting and to various glandular imbalances has likewise been cited as evidence of their greater physiological instability.

From these differences in physiological homeostasis, some writers have proposed a parallel sex difference in "mental homeostasis" (20). To this they attribute the greater "psychic unrest" of the female, as evidenced by more frequent emotionality, neurotic tendencies, nervous habits, feelings of inadequacy, and other symptoms of instability. The analogy is interesting, but we must proceed with the utmost caution in making such a transition from physiological to behavior data. Even in physical functions, exceptions can be found to the greater stability of the male sex. Moreover, we cannot *assume* that psychological and physiological homeostasis are necessarily related to a very high degree. It is true, for example, that physiological changes occur during emotional excitement, but it does not follow that individual differences in emotionality are correlated with individual differences in physiological characteristics. Furthermore, the physiological changes themselves may be influenced by the individual's previous experiences, home background, and the like. In fact, the evidence on individual differences in personality development tends to emphasize the role of experiential factors. Such factors may be equally important in determining sex differences in behavior.

THE ROLE OF CULTURAL FACTORS

That sex roles and sex stereotypes vary in different times and places is apparent not only from anthropology but from our own cultural history as well. To be sure, a few persistent differences in behavior can be identified. These undoubtedly result from some of the physical differences considered in the preceding section. Thus the widespread prevalence of male dominance in different cultures may be historically related to sex differences in physique and muscular strength. But the amount of such sex differences in dominance varies widely from culture to culture, as does the manner in which it is expressed. Moreover, many characteristics associated with the traditional male stereotype in our culture may be absent or reversed in other cultures (cf. 22).

Occupations have traditionally provided one of the principal cultural areas of sex differentiation. In relatively primitive cultures, in which occupations are predominantly physical, a sharp division of labor between the sexes is necessarily observed. Figure 90 presents a summary of the sex distribution of ten principal activities, based upon data from 224 representative tribes throughout the world. Because the

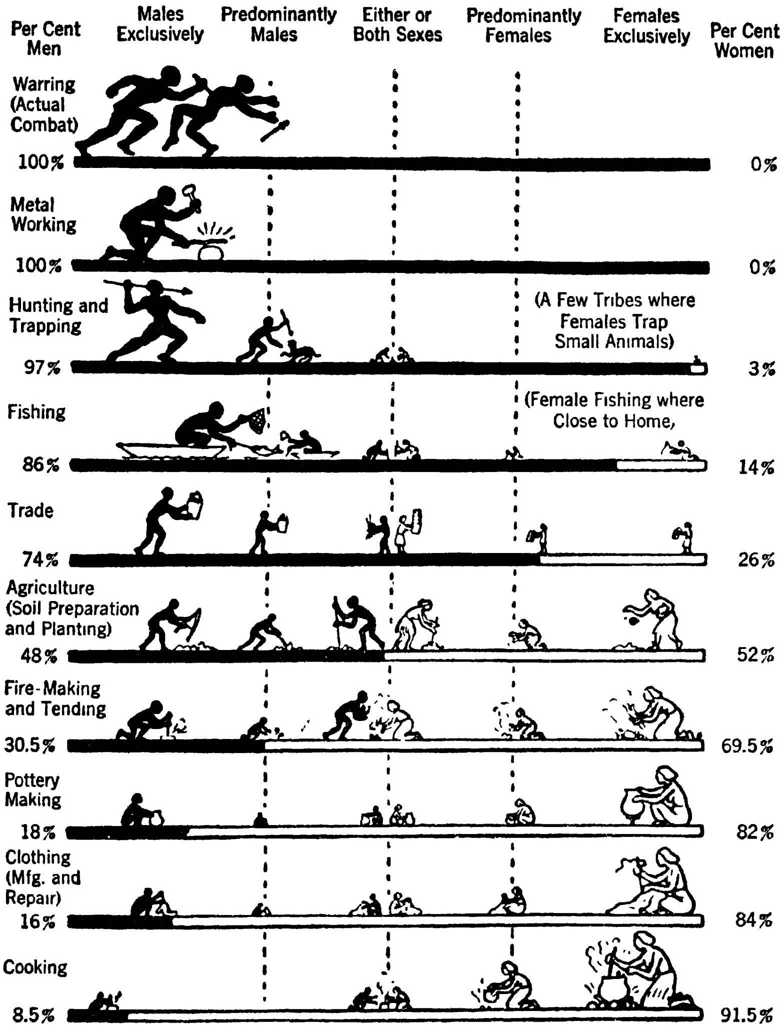


Fig. 90. Division of Labor Between the Sexes in 224 Representative Tribes Throughout the World. Size of figures shows the approximate degree to which each sex participates exclusively, predominantly, or together with the other sex. Heavy lines beneath figures indicate approximate total percent of male or female participation in each occupation, black space indicating male and white space female participation. (From Scheinfeld, 41, p. 293. Original data from Murdock, G. P., "Comparative Data on the Division of Labor by Sex," *Soc. Forces,* 1937, 15, 551–553.)

female bears and suckles the young, she is likely, in such primitive cultures, to engage in occupations which keep her closer to home, such as the preparation of food and the manufacturing and repair of clothing. Superior muscular strength and endurance cause the men to take over warring, metal work, hunting, and most of the fishing. But modern occupations do not fit into these primitive categories (41, 44). Even modern warfare is not so much a matter of handling spears and javelins as it is a matter of pushing buttons and designing blueprints. Paradoxically, it is the home that is now one of the principal loci of physical occupations, in contrast to the office, the store, the conference room, or the auditorium. With the development of machinery, the physical demands of more and more occupations are becoming reduced. Our thinking should not, therefore, be hampered by traditional stereotypes, but rather should be guided by the demands of the specific situation and the abilities of the specific individual. Several writers have called attention to the need for revising our conception of sex roles in terms of developments in modern living (32, 43).

That women have no "natural affinity" for certain tasks, nor men a "natural repugnance" toward their performance, can be amply illustrated. Huxley and Haddon (19, p. 69), in discussing the influence of social pressure upon sex differences in aptitudes, cite the remark of the third century Greek gossip writer, Athenaeus, "Whoever heard of a woman cook?" In the same vein, Mead (31, p. xix) calls attention to "the convention of one Philippine tribe that no man can keep a secret, the Manus assumption that only men enjoy playing with babies, the Toda prescription of almost all domestic work as too sacred for women, or the Arapesh insistence that women's heads are stronger than men's." Other illustrations can be found in the history of our own culture. Most writers on the social history of the Middle Ages, for example, call attention to the "masculine character" of women of that period. Thus Garreau, writing about France at the time of the crusades, has this to say:

> A trait peculiar to this epoch is the close resemblance between the manners of men and women. The rule that such and such feelings or acts are permitted to one sex and forbidden to the other was not fairly settled. Men had the right to dissolve in tears, and women that of talking without prudery. . . . If we look at their intellectual level, the women appear distinctly superior. They are more serious; more subtle. With them we do not seem dealing with the rude state of civilization that their husbands

belong to. . . . As a rule, women seem to have the habit of weighing their acts; of not yielding to momentary impressions.[14]

The play activities of boys and girls have been a subject of frequent discussion. Some would argue, for instance, that girls play with dolls because of a nascent "maternal drive" or some similar innate interest or emotional trait characteristic of their sex. The almost complete absence of this type of play activity among boys has accordingly been regarded as indicative of a fundamental biological diversification in emotional response. An observation made by Mead (30) in her studies on the island of Manus in New Guinea is of interest in this connection. Dolls are ordinarily unknown to the children on this island. But when they were presented for the first time with some wooden statuettes, it was the boys and not the girls who accepted them as dolls, crooning lullabies to them and displaying typical parental behavior. This reaction can be understood in terms of the pattern of adult behavior in Manus. Owing to the traditional division of labor, the women are busy with their various duties throughout the day, while the men have much more leisure time between their activities of hunting and fishing. As a result, the father rather than the mother attends to the children and plays with them. This socially established differentiation of behavior was reflected in the play responses of the boys and girls.

Another vivid illustration of the role of cultural factors in sex differences in behavior is furnished by a subsequent series of observations reported by Mead (31). These concerned the traditional emotional characteristics of men and women in three primitive societies in New Guinea. The three groups were sharply contrasted in the pattern of male and female personality which they presented. Among the *Arapesh*, both men and women displayed emotional characteristics which in our society would be labeled distinctly feminine. In this group both sexes are trained to be cooperative, unaggressive, gentle, non-competitive, and responsive to the needs of others. They are strongly imbued with a sense of obligation toward any who are weaker or younger than themselves. Even their typical response toward material objects is not one of possession but of solicitude.

The *Mundugumur*, a river-dwelling tribe of cannibals and head-hunters, present a sharply contrasting picture. In this society both men

[14] Garreau, L. *L'état social de la France au temps des Croisades.* Paris, 1899. Quoted in 1, p. 199.

and women are violent, aggressive, ruthless, and competitive. They take great delight in action and in fighting. They are quick to perceive an insult and ever ready to avenge it. Because of an intricate system of family organization, the child is born into a hostile world, in which most members of his own sex are his enemies. This is particularly true of boys, but a child of either sex will be disliked and resented by some members of the family.

Perhaps the most interesting pattern is presented by the *Tchambuli,* among whom there is a genuine reversal of the sex-attitude of our culture. It is the women who have the position of power in Tchambuli. The group depends for its food supply upon the fishing of the women, the men rarely engaging in this activity. Fish is also the staple product of trade, in exchange for which several essential commodities are obtained. Similarly, it is the women who make mosquito bags, the most important article of Tchambuli manufacture and in great demand by outside purchasers. The men, on the other hand, engage predominantly in artistic and non-utilitarian pursuits. Most men are highly skilled in more than one art, including dancing, carving, painting, and others. It is the man in this society who is concerned with the beauty and elaboration of his costumes and the excellence of his artistic accomplishments. This type of life is reflected in pronounced personality differences between the sexes. The women are impersonal, practical, and efficient. Their attitude toward the men is one of kindly tolerance and appreciation. The men are graceful, artistic, emotionally subservient, timid, sensitive to the opinions of others, and throughout their lives dependent upon the security afforded to them by the women.

As in our society, each of these three cultures has its "deviants," its maladjusted individuals whose personality traits clash with the accepted standards. But the deviant in one society often coincides with the traditional ideal of another. Thus the "masculine" woman among the Tchambuli is one who embodies the typically feminine characteristics of our society; the "effeminate" Tchambuli man displays behavior which we would characterize as typically masculine. In a final evaluation of her findings, Mead writes:

> We are forced to conclude that human nature is almost unbelievably malleable, responding accurately and contrastingly to contrasting cultural conditions. The differences between individuals who are members of different cultures, like the differences between individuals within a culture, are almost entirely to be laid to differences in conditioning, especially

during early childhood, and the form of this conditioning is culturally determined Standardized personality differences between the sexes are of this order, cultural creations to which each generation, male and female, is trained to conform (31, pp 280–281).

It is apparent that cultural factors play an important part in the differentiation of sex roles and in the corresponding sex differences in behavior Moreover, even when physical differences contribute to sex differences in behavior, the contribution is usually indirect and intricately overlaid with cultural factors In such cases, it is the *social implications* of such physical differences, rather than the biological sex differences themselves, which lead to divergent personality development in the two sexes.

CHAPTER

19

Sex Differences: Major Results

THE DATA WHICH WILL BE SURVEYED in the present chapter concern *sex differences under existing conditions in our society.* Such data, although limited in their application, are not without value. Thus it is of considerable *practical* interest to ascertain the typical behavioral characteristics of men and women, whatever may be the origin of the differences. The number of situations in which such knowledge may prove useful is legion. In many fields of activity, definite assumptions are made in regard to existing sex differences in aptitudes, interests, emotional responses, and similar traits. This sex differentiation is noticeable in advertising and selling, job placement, political campaigning, the organization of newspapers and magazines, social work, crime prevention, and the treatment of offenders, to name only a few outstanding examples. In a descriptive account of any one cultural group, the question of sex differences in behavior can be legitimately raised. Regardless of whether such differences are the indirect result of structural dissimilarities or whether they have an exclusively cultural or environmental origin, they cannot be ignored in the practical adjustments of everyday life.

It is also possible that a careful analysis of the material on sex differences, in conjunction with other available information, may help to clarify the nature and source of such differences. Such an approach can never furnish a conclusive account of the origin of sex differences, but it may indirectly yield some corroborative evidence on this problem.

In view of the problems discussed in the preceding chapter, such as selective factors, extensive overlapping of groups, errors of sampling, errors of measurement arising from inadequacy of the tests, and unwarranted generalizations regarding the functions measured, it is obvi-

ously difficult to formulate any summary statements regarding sex differences from the data of a number of independent investigations. This is especially true since such investigations differ widely in number and kind of subjects, specific tests or materials employed, and other important conditions. Similarly, all but the most recent and best controlled studies fail to report reliabilities of differences, degree of overlapping, and other essential facts, thus making it difficult to evaluate their findings. In the face of these conditions, the only available criterion for the acceptance of a conclusion is the *consistency of results of different investigators.* A survey of the experimental literature on sex differences reveals certain major findings which are so frequently reported by different investigators as to suggest a valid basis in fact. It is with these findings that we shall be primarily concerned.[1]

SIMPLE SENSORI-MOTOR FUNCTIONS

In *sensory acuity,* sex differences are slight and inconsistent, with the exception of the female superiority in color discrimination. Color-blindness is found in about eight times as many men as women; the most common, or red-green, form of color-blindness occurs in about 4% of the general male population and only about 0.5% of the female. There is fairly conclusive evidence, moreover, that the most common form of color-blindness is a sex-linked hereditary deficiency (cf. Ch. 4). Even among individuals of normal vision, females excel in color discrimination. For adults, such a difference could be attributed to the greater amount of practice which women have had in the use of color, as in matters of dress, embroidery and other needlecrafts, interior decoration, and the like. The female superiority in color discrimination has, however, been observed even in early infancy (80). These results raise an interesting point which may also be related to the explanation of certain other sex differences in behavior. Owing to sex differences in rate of maturation (Ch. 18), it is possible that female infants excel male infants of the same chronological age in their color responses simply because the females are farther along in their development. The fact that girls have a "head start" in this func-

[1] No attempt will be made to present a survey of specific studies on sex differences in behavior. Such material has been periodically reviewed by various writers. Cf. Allen (2), Goodenough (28), L. S. Hollingworth (39, 40), Lincoln (51), Louttit (54), C. C. Miles (60), Wellman (94), Woolley (87, 100, 101), and the more recent reviews by Johnson and Terman (42) and Terman *et al.* (86).

tion may in turn tend to perpetuate their advantage, through greater interest and experience in handling colors. In other sensory modalities, such as taste, smell, hearing, and touch, the data either show no significant sex difference or are difficult to interpret because of the presence of uncontrolled factors.

In tasks involving the rapid *perception of details* and frequent shifts of attention, women generally excel. This is one of the principal abilities measured by clerical aptitude tests, on which women make a consistently better showing than men. In the norms reported for the Minnesota Clerical Test, only about 16% of male workers in the general population reached or exceeded the median of female workers in checking similarities or differences in lists of names and numbers (65). Moreover, a series of different investigations showed a significant female superiority on this test from the fifth grade through the senior year of high school (65, 74).

Fairly large and consistent sex differences have been reported in various aspects of *motor performance*. On the average, boys surpass girls not only in muscular strength, but also in speed and coordination of gross bodily movements. This difference has been observed from infancy. In extensive observations of children of preschool age, Gesell and his co-workers (25) found that boys were faster and made fewer errors in walking a series of narrow boards. The boys also achieved more accuracy and greater distance in throwing a ball than did girls of the same age. In connection with the latter observation, a study was also made of the characteristic ball-throwing pattern of boys and girls. A clear-cut sex differentiation in the typical ball-throwing stance was already apparent among 5- and 6-year-olds. Males of all ages average better than females on such coordination tests as aiming and tracing. Men have also been found to have shorter and more consistent reaction times than women.

In manual dexterity, on the other hand, girls generally excel (25). In early childhood this is exemplified by the fact that girls are usually able to dress themselves at an earlier age and more efficiently than boys. Girls' superior control of finger and wrist movements is also indicated in such behavior as hand washing and turning door knobs. In the standardization of the 1937 Stanford-Binet, more girls than boys passed the tests on buttoning and on tying a bowknot. The statistical significance of these sex differences was exceptionally high (59). That adult women can perform many manipulatory tasks more

quickly and accurately than men has been widely recognized in industry. This fact was especially apparent during World War II, when women were frequently assigned to assembly, inspection, and similar industrial operations. Such an observation is also supported by aptitude test performance. On tests like the O'Connor Finger Dexterity Test, O'Connor Tweezer Dexterity Test (cf. 8), and Purdue Pegboard (90), the norms for adult women are consistently higher than those for men.

The male superiority in gross bodily movements may be largely the result of such structural factors as muscular strength and bodily size and proportions. The female advantage in manual dexterity and speed and control of fine movements, on the other hand, may arise initially from the developmental acceleration of girls. In general, delicate movement follows gross bodily movement within the development of the individual. Girls would thus be expected to develop fine motor coordinations at an earlier age than boys. These initial, structurally determined sex differences may affect the acquisition of interests and skills, thereby setting in motion a progressive mechanism of differentiation between the sexes.

It has been suggested, for example, that the differences in motor development may help to explain why girls play with dolls much more commonly than boys.[2] The detailed hand movements involved in dressing and undressing dolls and in related play activities may appeal more to girls because of their superior manual dexterity. At the same time, it would be legitimate to ask, "Why *dolls*?" On the basis of manual dexterity alone, many other types of toys would qualify, including erector sets, mechanical puzzles, and toy clocks which could be taken apart and re-assembled. But girls in our society are not generally given such toys. Moreover, they usually receive their first doll when they are too young to do anything but throw it across the room, in typically "masculine" fashion. This is but one illustration of the fact that cultural influences are so inextricably involved in behavior, from its earliest beginnings, that it is well-nigh impossible to trace the effect of structural factors *per se*.

INTELLECTUAL FUNCTIONS

General Intelligence. On most intelligence tests of the common verbal type, sex differences are slight, but more often in favor of girls

[2] Dr. Helen Thompson, quoted in Scheinfeld (73), p. 92.

than boys (44, 86). On such widely used group tests as the National Intelligence Test, for example, girls excel consistently (44, 70). Selective factors in sampling sometimes produce misleading results. Thus in high school groups, boys generally obtain higher averages than girls, since the duller boys tend to drop out of school in larger numbers than the duller girls (Ch. 18). Similarly, the exclusion of institutionalized cases or of children in special classes tends to favor the boys, since a larger proportion of girls of correspondingly low intelligence remain in regular classes (Ch. 18). When samplings are fairly comparable, however, most intelligence tests show either no significant sex difference or a slight difference in favor of girls.

The female advantage on many intelligence tests has been found from early childhood to late maturity. In one study (29) on preschool children, the Kuhlmann-Binet was given to 50 boys and 50 girls at each of the ages 2, 3, and 4. The average IQ of the girls was higher than that of the boys, the difference persisting when the children were retested after a lapse of six weeks. Thus the obtained difference could not be attributed to chance fluctuations or to a sex difference in response to novel situations. A similar difference was found in the study of mental growth and decline conducted in rural New England communities and described in Chapter 9 (16). Within the entire sample of 581 men and 607 women between the ages of 10 and 60, the women obtained a significantly higher average on the Army Alpha. In some of the separate age groups, the sex difference was insignificant, although still favoring the women.

Surveys with the Stanford-Binet have tended to show negligible sex differences. In the Scottish survey discussed earlier (Chs. 3 and 18), all children born in Scotland on each of four specified days were given the 1916 Stanford-Binet. The 444 boys thus tested had an average IQ of 100.5 and the 430 girls 99.7. The critical ratio of this difference is only .86, indicating that the difference is no greater than one would expect from any two samples of the same sex. A fundamental point to consider in this connection is that in a carefully standardized individual test such as the Stanford-Binet, items which favor one sex or the other are deliberately omitted. This is particularly true of the 1937 revision of the Stanford-Binet (cf. 59). Items which showed a large sex difference in the percentage passing were excluded entirely, on the assumption that sex differences on such items may be specific to the task in question and may simply reflect differences in experience and

training. Among the remaining items, those slightly favoring girls were balanced against others which favored boys to an equal degree. The fact that no significant sex difference in IQ was found in the standardization sample of the 1937 Stanford-Binet is therefore an index of the care with which this procedure was followed, and has little or no bearing upon sex differences in intelligence.

Whether boys or girls obtain higher IQ's depends upon the items which are included in the test. When no deliberate effort has been made to exclude sex differences from the test, there has generally been a tendency to favor girls. This follows from the fact that intelligence tests consist so largely of verbal items, on which girls are superior.[3] In so far as the tests depend upon memory, girls have an additional advantage.[3] Moreover, many intelligence tests are validated against school achievement, in which girls also excel, especially at the elementary school level. It is apparent from this discussion that the question of which is the more "intelligent" sex is somewhat ambiguous. In the light of what we now know about trait organization and the nature of intelligence (Ch. 15), this is not surprising. It is much more meaningful to ask what sex differences exist in the more specific functions which make up "intelligence" in our culture.

Special Aptitudes. Female superiority in *verbal* or linguistic functions has been noted from infancy to adulthood (58, 86). This difference is found in almost every aspect of language development which has been studied, and has been reported with remarkable consistency by different investigators. In fact, the few results which fail to support this difference—or, more rarely, reverse it—can usually be explained either by selective factors or by the use of material which appeals much more to the interest of boys (cf. 58). Observations on normal as well as on gifted and feebleminded children have shown that on the average girls begin to talk earlier than boys. Similarly, girls of preschool age have a larger vocabulary than boys. In one study (57), the percentage of comprehensible verbal responses was determined for each child. At 18 months, the average per cent was 14 for boys and 38 for girls; at 24 months it was 49 for boys and 78 for girls. Girls likewise begin to use sentences earlier than boys and tend to use more words in sentences. In learning to read, girls make more rapid progress than boys (72, 98).

[3] The evidence for these sex differences in specific intellectual functions will be discussed in the following section.

Girls also reach maturity in articulation at an earlier age than boys. The articulatory patterns of girls in the first school grade are approximately the same as those of boys in the second grade. This developmental difference in the motor aspects of speech may provide a clue to the general female superiority in linguistic functions. The acceleration of girls in physical development probably accounts for their more rapid progress in articulation. This in turn may give them a powerful initial advantage in the mastery of all phases of language. Such a difference in developmental rate may also account in part for the much greater frequency of reading disabilities, stuttering, stammering, and other speech disorders among boys. The ratio of male to female stutterers varies from 2:1 to 10:1 (75, 76). In a survey of 17 groups of reading disability cases (5), the proportion of boys varied from 60% to 100%. If, in speaking and reading, boys are more often held up to standards which they are not structurally ready to meet, they may experience more frustration, loss of confidence, and confusion in linguistic situations than girls (58, 75, 76). This may be an important factor not only in the development of linguistic disorders but also in the normal individual's subsequent progress in verbal functions.

The verbal superiority of girls persists throughout the successive educational levels, the sex difference often becoming more pronounced at the upper levels. Girls usually excel in speed of reading and in such tests as opposites, analogies, sentence completion, and story completion. In a study of the language development of children in grades 4 to 12, 472 boys and 514 girls were asked to write a composition on a prescribed topic of interest to both sexes. Within the same time limit, girls produced longer themes than boys, the elementary school boys using on the average 86% as many words as the girls, and the high school boys 83% (45).

A mass of relevant data is also provided by the analysis of sub-test performance on intelligence tests. In the standardization sampling of the 1937 Stanford-Binet (59), a significantly greater percentage of girls passed some of the sentence completion and code learning tests.[4] In the two investigations with the Pressey Group Test of Intelligence reported in the preceding chapter, an analysis of sub-test scores showed fairly consistent sex differences in the elementary school and high school samplings (9, 69). In both samples, the girls excelled in word

[4] The specific tests are Minkus Completion (Form L, Year Level XII, Test 6) and Codes I (Form M, Average Adult Level, Test 4).

completion and dissected sentences, and in the elementary school group they also surpassed the boys in opposites and analogies. It will be recalled that among the high school seniors, owing to the differential elimination of male and female students, boys excelled in total score on this test, whereas girls had excelled at the elementary school level. Performance of boys and girls on the separate tests, however, showed the same *relative* standing in both educational groups. The reversal in total score from the elementary to the high school groups resulted from the fact that the high school senior boys excelled *by a much larger amount* in the same tests in which the elementary school boys excelled. Similarly, the high school senior girls excelled *by a smaller amount* in the tests in which the elementary school girls had excelled markedly.

Large and significant sex differences in verbal functions have also been found on the psychological tests administered to entering college freshmen. Table 37 shows the average scores of several thousand en-

TABLE 37 *Sex Differences among College Freshmen in the Verbal and Mathematical Sections of the Scholastic Aptitude Test*

(Adapted from Brigham, 10, p. 383)

	Number of Cases	*Average Score*	
		Verbal	Mathematical
Boys	4214	486.58	511.15
Girls	3362	512.29	476.74
diff./$\sigma_{diff.}$		11.34	15.27

tering students of each sex on the verbal and numerical parts of the Scholastic Aptitude Test administered in 1931 by the College Entrance Examination Board. The difference in favor of the girls on the verbal section is over eleven times as large as its standard error, considerably larger than the critical ratio of 2.58 required at the .01 level of confidence.[5] On the American Council Psychological Examination

[5] In the samplings tested during the later 1930's and the decade of the 1940's, this sex difference tended to disappear, the boys' and girls' averages in the verbal score being practically identical. Changes in college admission policies as they affected the two sexes, as well as other selective factors, may account in part for this shift in relative standing. In all years, however, the girls did much better on the verbal than on the numerical parts of the test, while the reverse was true of the boys. Cf. College Entrance Examination Board. *Annual Reports of the Director, 1935 1948.*

(ACE), sex differences in *total* L-score, based upon the three verbal sub-tests, are negligible and inconsistent from year to year. Some of these linguistic sub-tests may, however, depend to a considerable extent upon general information, in which boys consistently excel. Studies on Negro college students (13), as well as on white, Chinese, Japanese, and part-Hawaiian high school graduates tested in Hawaii (52), showed a highly significant female superiority in the artificial language sub-test [6] of the ACE. This difference appeared consistently and reliably in all sub-groups. The artificial language test attempts to measure the ability to handle linguistic relations, independently of the individual's general information. Given a short vocabulary and a few simple grammatical rules, the subject is required to "translate" a short English passage into this artificial language.

Girls also excel in most tests of *memory*, although the differences are neither so large nor so consistent in this respect as they are on verbal tests. In the standardization sample of the 1937 Stanford-Binet (59), a significantly greater percentage of girls passed the tests of picture memories and copying a bead chain from memory. No significant difference in favor of either sex was found on other memory tests in the scale.[7] Group tests of intelligence also tend to show superior female performance on sub-tests involving memory (9, 69, 86). In digit span and in memory for geometric forms, however, sex differences are negligible and inconsistent (86). In memory for narratives, the direction of sex differences often depends upon the relative appeal of the content for the two sexes.

In general, however, when the content favors neither sex, girls tend to excel more consistently in logical than in rote memory. This may result from the greater dependence of logical memory tests upon verbal comprehension. It is possible, in fact, that the female superiority in many memory tests is attributable to the role of verbal functions in facilitating retention and recall of most types of material. Another relevant observation is that women seem to have more vivid mental imagery than men in every sense modality. This finding, first suggested by Galton on the basis of his famous "breakfast table" questionnaire (24), has been subsequently corroborated by several

[6] This test was not retained in the more recent forms of the ACE.

[7] With the exception of memory for a story about acrobats, which obviously appealed more to the interests of boys than to those of girls.

investigators. To what extent such a difference may be the result of sex differences in occupations and other traditional activities remains to be seen.

A difference in favor of the male sex has been repeatedly observed in various phases of *spatial and mechanical aptitude*. The possibility that this difference has a predominantly cultural basis, however, is suggested by several facts. Thus male superiority is more pronounced and consistent in tests depending upon mechanical information than in the more abstract tests of spatial relations, which may be equally unfamiliar to both sexes. Moreover, male superiority in this area is not evidenced as early as was the female superiority in verbal aptitude. For example, in the extensive observations by Gesell and his associates at Yale, no significant or consistent sex differences were found during the first five years of life in tests involving block building, form boards, and form recognition (25). On the other hand, in a study of 100 children between the ages of 2 and 4 with the Wallin Peg Board, Goodenough (29) reports a difference in favor of boys. In this test, the subject is required to insert round, square, and triangular pegs into the appropriately shaped holes as rapidly as possible. The boys obtained a significantly higher average on this test, despite the fact that in the same group the girls excelled significantly in Kuhlmann-Binet IQ. In the light of the negative findings in the Gesell observations, it is likely that the sex differences reported on such isolated tests may result from differences in the play experiences of the particular groups of boys and girls studied.

Among children of school age, a clear-cut sex differentiation on mechanical tests is already apparent. On the Stanford-Binet, boys were found to excel significantly in block counting from pictures, directional orientation, and plan of search, all of which probably involve spatial abilities (59). On such tests as form boards, puzzle boxes, assembling objects, and slot mazes, boys also score much higher than girls in both speed and accuracy. A similar male superiority was found by Porteus (68) in his graded paper-and-pencil mazes. Boys clearly excelled on these mazes, when compared with girls of the same Stanford-Binet IQ's.

Interesting sex differences at two age and educational levels were reported in connection with the standardization of the Minnesota Mechanical Aptitude Tests (66). Seventh grade boys and girls, as well as college sophomores, were employed in these comparisons. In

Table 38 will be found the critical ratios of the differences between male and female averages on each of the tests in the battery. The largest and most consistent sex difference is noted on the Assembly Test, which requires the assembling of a number of common objects, such as a bottle stopper or a spark plug, from the given parts. The greater experience of boys with mechanical objects undoubtedly gives them an advantage on such a test. The Paper Form Board Test, involving more abstract spatial visualization, shows a male superiority which falls short of that required at the .01 level of confidence. The Spatial Relations Test calls for the insertion of numerous irregularly shaped pieces in their appropriate recesses as rapidly as possible. This test, together with Block Packing and Card Sorting, favors girls in accordance with the commonly reported sex difference in manual dexterity and perceptual discrimination. It should be noted, however, that the female advantage on these tests disappears in the college sampling, either because of selective factors or because of intervening experiential differences.

TABLE 38 *Critical Ratios of the Differences between Male and Female Averages on the Minnesota Mechanical Aptitude Tests*

(From Paterson *et al.*, 66, p. 274)

Test	*Critical Ratio* (*diff*/σ_{diff})*	
	Seventh Grade Pupils	College Sophomores
Assembly	12.1	10.4
Paper Form Board	2.0	2.4
Spatial Relations	−3.2	2.4
Block Packing	−5.0	1.4
Card Sorting	−8.9	−0.6

* In this table, a minus sign indicates a difference in favor of the girls.

On tests of mechanical comprehension, women score lower than men, as would be expected on the basis of sex differences in mechanical information and experience. This is illustrated by the results of a survey with the Bennett Test of Mechanical Comprehension (Form AA), given to 390 females and 338 males of comparable age and education, including high school and adult groups (6). The males averaged much higher than the females on this test, the critical ratios of the differences ranging from 7.2 to 10.5 in different groups. Al-

though the sex difference varied considerably from problem to problem, the females as a group made more errors than the males on every item.

Mention may also be made of sex differences on performance tests of "intelligence," many of which depend largely upon spatial rather than verbal aptitudes. Such tests generally favor the boys. For example, in the complete sampling of Scottish children previously described, a battery of eight performance tests, selected from well-known intelligence scales, was administered in addition to the Stanford-Binet (55). The total performance score showed a significant difference in favor of the boys, the critical ratio of the differences being 3.74.

On *numerical* tests, the largest differences are again in favor of boys. Such a male advantage fails to appear, however, until the children are well into the elementary school period. Gesell's observations on preschool children show either negligible sex differences or a slight superiority of girls in the early development of numerical concepts (25). Extensive surveys on kindergarten and first grade children have also yielded no significant sex difference in arithmetic abilities (11, 99). At the lower levels of the Stanford-Binet, sex differences on tests involving counting and number concepts are likewise negligible or inconsistent (86).

Among elementary school children as well as older subjects, computation tests show either no sex difference or, more often, a difference in favor of girls (86). On arithmetic problems and other numerical reasoning tests, males excel quite consistently (86). In the 1937 Stanford-Binet (59), boys excel significantly on the tests of arithmetic reasoning, ingenuity (a more difficult type of numerical reasoning problem), and induction (in which a generalized numerical rule must be found). In the previously cited studies with the Pressey Group Test of Intelligence (9, 69), the boys excelled on the arithmetic reasoning test at each age in the elementary school group, as well as in the high school senior group. On the Army Alpha given to 834 high school students, the boys excelled significantly in only three tests: arithmetic reasoning, number series completion, and information (95). The differences on these three tests were sufficient to pull up the total scores and produce a difference in favor of the boys on the scale as a whole.

The scores of college freshmen reproduced in Table 37 show a highly significant difference in favor of males in the mathematical part

of the Scholastic Aptitude Test (10). The critical ratio of this difference is over 15, indicating virtually complete certainty that the difference could not have arisen from sampling fluctuations. Similar differences in favor of males were found in surveys with the American Council Psychological Examination (ACE) on American Negro students (13) and on white, Chinese, Japanese, and part-Hawaiian high school graduates tested in Hawaii (52). In all these groups, the arithmetic test yielded significant and consistent differences in favor of the males. The annual ACE norms for American colleges also show a significant difference in favor of males in *total* Q-score, based upon arithmetic reasoning, number series completion, and figure analogies. The first two tests are numerical and the third spatial in content. In the 1947 norms, based on 30,924 males and 24,918 females, the mean Q-score was 44.39 for males and 41.40 for females. The critical ratio [8] of this difference is 30.92.

An interesting recent development in the study of sex differences has involved the comparison of boys and girls on the Chicago tests of *Primary Mental Abilities,* designed by Thurstone on the basis of factor analysis (cf. Ch. 15). In one survey (38) on eighth and ninth grade pupils, the girls excelled significantly in word fluency, reasoning, and visual memory, while the boys were significantly superior in spatial orientation. No other sex differences were found to be clearly and consistently significant at both grade levels. In another study (35), the same tests were given to all 13-year-olds in a small midwestern town. This group included a total of 40 boys and 51 girls, and ranged from the fourth to the ninth school grades, although most of the children were in the eighth grade. The spatial test was again the only test on which a difference approaching statistical significance favored the boys. The girls were significantly superior on the reasoning and number tests. They also excelled in word fluency and visual memory, but the significance of these differences was low. The two sexes were practically equal on the verbal comprehension test.

It is probably premature to draw any conclusions regarding sex differences on these tests of "primary mental abilities," but on the whole the findings corroborate those obtained with other tests. The boys are significantly superior on the spatial aptitude tests. The girls excel consistently in word fluency and visual memory The number test, which also favors the girls, consists exclusively of arithmetic com-

[8] Computed by the writers from data given in 89, p. 14.

putation, in which girls have previously been found to excel. Similarly, female superiority in the reasoning test may possibly be related to the fact that the test consists of letter series to be completed. Girls may have more facility with this type of material, in connection with their general linguistic superiority. On the other hand, the girls' failure to excel in the verbal comprehension test may be due to the contribution of general information to performance on this test. Since boys usually excel in general information, this could counteract the girls' superiority in verbal ability.

Finally, we may consider briefly certain sex differences in *artistic* abilities. Among preschool children, girls generally include more details in their drawings than do boys (25). This is true in their spontaneous drawings as well as in various controlled drawing tests, such as drawing a man or completing figures. One reason for such a difference may be the possibility that girls are more observant of details, another that they spend more time in drawing during early childhood. Both of these differences could in turn result from the fact that the activities of girls are traditionally more sedentary and more circumscribed than those of boys. Girls would thus be more likely to notice minute details in their surroundings and would also have more practice in such relatively sedentary pursuits as drawing. In later childhood and adulthood, sex differences in artistic production or appreciation are even more difficult to evaluate because of obvious differences in relevant training and experience. On such tests of art appreciation as the McAdory and the Meier Art Judgment Test, women exceed men in average scores by small but fairly significant amounts (20).

In the Seashore Tests of Musical Talent, which measure relatively simple auditory discrimination and memory, no significant sex differences have been found (21). On more complex tests, placing greater emphasis upon aesthetic appreciation, the scores generally favor women. An interesting clue to the probable origin of some of these differences is provided by an investigation on college students with the Kwalwasser-Dykema music tests (26). Comparisons were made between men and women within a group of 1000 students in twelve eastern colleges. In the total undifferentiated groups, women excelled in average score. But when only subjects who had received no musical training were compared, the sex difference disappeared. These findings thus suggest that the sex differences ordinarily reported

on such tests may result from differential amounts of training received by the two sexes.

SCHOOL ACHIEVEMENT

On the whole, girls excel in *general school achievement*, as revealed both by achievement test results and by school grades. Performance on the separate parts of standardized *achievement tests*, however, shows a hierarchy of abilities in different school subjects which corresponds closely to that found with tests of intelligence and special aptitudes. Corresponding sex differences have been found in the expressed *preferences* for different school subjects among elementary and high school students (86, pp. 964–966). The same hierarchy of academic achievement has been reported consistently from the elementary school (37) to high school (43) and college (79), and from morons (67) to gifted children (85). On such tests as the Stanford Achievement battery, given to many thousands of elementary school children in a number of independent surveys, the boys excelled in arithmetic reasoning, nature study, science, and history; the girls, in reading, language usage, spelling, and arithmetic computation. These sex differences in achievement persist when boys and girls are equated in general intelligence. For example, in an investigation on high school students, 410 boys and 349 girls were given both the Terman Group Test of Intelligence and a standardized geometry test (93). The average geometry test score of the boys was higher than that of the girls, when the two sexes were equated in intelligence test score.

Some of the largest sex differences in achievement test scores have been reported on science tests. In a survey (43) of North Carolina high school seniors, including approximately 8000 boys and 11,000 girls, the critical ratio of the difference in favor of the boys was 31.7 on the science section of the achievement test employed.[9] Similarly, in a college survey (48) on 2992 men and 1410 women, the mean difference in natural science scores was about 24 times as large as its standard error. In the annual Science Talent Search, sponsored by Westinghouse, boys again averaged higher than girls (18).

In summary, girls surpass boys in those school subjects depending largely upon verbal ability, memory, and perceptual speed. Boys excel

[9] One of the reasons for this exceptionally large critical ratio is, of course, the size of the samplings.

in those subjects which call into play numerical reasoning and spatial aptitudes, as well as in certain "information" subjects such as history, geography, and general science. This is in agreement with the common superiority of boys on tests of general information included in intelligence scales, and probably results from the less restricted and more heterogeneous environment to which boys are exposed, as well as from their wider range of reading interests. Terman (85), for example, in his survey of the reading habits of gifted children, reports that the girls read imaginative and emotional fiction as well as stories of school and home life far more often than the boys, while the latter showed a predominant interest in books on science, history, biography, travel, and informational fiction and adventure tales.

In regard to *school progress*, girls are consistently more successful than boys. The differences, although small, appear irrespective of the particular criterion of school progress employed (51, 53, 71). Girls are less frequently retarded, more frequently accelerated, and promoted in larger numbers than boys. Typical results from a survey conducted in the schools of 318 cities are shown in Table 39. Since girls make more rapid progress than boys in school promotions and

TABLE 39 *Median Percentage of Boys and Girls in Normal Age-Grade Location, as Well as Those over Age and under Age*

(From Lincoln, 51, p. 100)

School Status	*Cities of over 25,000*		*Cities of less than 25,000*	
	Boys	Girls	Boys	Girls
Normal	56	60	54	58
1 year over age	20	18	20	18
2 years over age	10	9	11	8
3 years over age	5	3	4	3
4 years over age	2	1	2	1
Total over age	38	32	38	36
Total under age	4	4	4	5

since most comparisons of achievement are made on the basis of school grade, it follows that age comparisons would show an even greater superiority of girls.

In *school grades* girls excel consistently, *even in those subjects which favor boys*. Thus a comparison of grades in arithmetic, or his-

tory, or any other subject in which boys obtain higher achievement test scores, shows a sex difference in favor of girls. The advantage enjoyed by girls in school grades was made particularly vivid in an investigation (50) on 202 boys and 188 girls in grades 2 to 6, all of whom were given the Stanford Achievement Test. The girls were found to excel consistently in school grades, *when compared with boys receiving the same achievement test scores*. Thus the grades showed a far greater female superiority than seemed to be warranted by performance on objective achievement tests.

Similarly, high school girls generally obtain better grades than high school boys, even though the latter are a more select group and make a better showing on achievement tests (cf. 53). This is illustrated by a survey of the grades given to students in each of the four years of a single high school, the results of which are shown in Table 40. It will be noted that in each year, without a single exception, the percentage of A's and B's is larger and the percentage of D's and F's is smaller among the girls than among the boys. The larger percentage of boys than girls who left school further suggests the better adjustment of girls to the school situation.

TABLE 40 *The Percentage of Each Letter Grade Received by Boys and Girls in a Single High School*

(From Lincoln, 71, p. 93)

Grades	*First Year*		*Second Year*		*Third Year*		*Fourth Year*	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
A	3.2	8.7	3.5	7.6	4.0	10.9	11.6	15.5
B	10.3	18.4	12.9	20.9	13.0	25.6	16.3	31.9
C	20.8	20.0	16.9	22.3	22.8	27.5	31.0	29.7
D	23.8	21.1	27.0	20.3	31.3	21.3	29.4	16.5
F	25.7	18.3	22.7	15.3	17.2	5.7	7.9	1.6
Left school	16.1	13.5	16.9	13.6	11.7	9.0	3.9	4.6

Various explanations have been offered for the greater academic success of girls. Among the major factors may be mentioned girls' demonstrated superiority in *linguistic aptitude*, which probably plays an important part in nearly all school subjects. Current methods of instruction, as well as methods of testing, are predominantly verbal. The child who expresses himself well, furthermore, will impress the

teacher as being relatively brighter than one who is linguistically backward, and this may in turn affect their respective grades. Another possible factor in the higher academic ratings of girls is the neatness and general superiority of their *handwriting*. In most investigations on both elementary and high school groups, girls have been found to excel markedly in the quality of their handwriting, as judged by standardized product scales (cf. 51, pp. 72–77). Such a difference may well affect the grades on school examinations as well as on written assignments.

Owing to the obvious presence of a subjective element in school grades, it is probable that *personality differences* between boys and girls also influence the allotment of such grades. The importance of this factor has been emphasized by several investigators (53, 71). Girls are generally more docile, quieter, not so subject to out-of-school distractions, less resistant to school discipline, and are less often "behavior problems" than boys. This difference in the child's attitude toward school affects his grades both through the amount of material actually learned and, more directly, through the general impression created on the teacher.

INTERESTS, PREFERENCES, AND ATTITUDES

That definite personality differences exist between adult men and women in our society is clearly apparent from everyday observation. In many emotional and social characteristics, this differentiation is noticeable from an early age. An important aspect of personality development in which traditional sex differences are manifested includes interests, preferences, ideals, attitudes, and personal sense of values. These characteristics, because of their relatively subtle and persistent nature, often exert an unsuspected influence, not only upon the development of emotional and character traits, but also upon the individual's achievements and effective abilities.

Data on sex differences in interests and attitudes are available from a wide variety of sources. Especially plentiful is the information gathered on children (cf. 86). The preferences of boys and girls have been compared in such areas as play activities, spontaneous drawings, the choice of topics for written compositions, collections, reading, movies, radio programs, favorite characters in fiction or in public life, vocational choices, and general life goals. Fairly clear-cut and

consistent male and female interest patterns have emerged from these varied studies.

A few typical investigations will serve to illustrate these findings. In their study of the *play activities* of 554 gifted and 474 unselected children, Terman *et al.* (85) computed a "masculinity index" for each of 90 common plays, games, and activities of childhood. For each activity, this index was based upon the relative knowledge, interest, and participation of boys and girls. Among the most "masculine" activities in the scale are listed: tools, shooting, kites, bicycling, marbles, wrestling, boxing, football, tops, machinery, baseball, and fishing. At the extreme of "feminine" activity are: dolls, dressing up, hopscotch, cooking, playing house, playing school, knitting or crocheting, dancing, sewing, playing store. In what is probably the most extensive collection of data on children's play activities, Lehman and Witty (49) questioned approximately 17,000 urban and 2000 rural children. In general, they found that boys engage more often in active vigorous play, in activities involving muscular dexterity and skill, and in highly organized and competitive games. The play of girls tended to be more sedentary, conservative, and restrained in range of action. Observations on children in kindergarten and the primary grades have shown that boys devote much more time to playing with building material, girls to painting and modeling (22).

A number of surveys have been conducted on sex differences in expressed *reading preferences*, as well as in the books actually used in libraries. In order to avoid the possible effects of the child's reading skill and of his previous familiarity with certain books, some investigators have employed lists of fictitious book titles. In one such study (88), about 200 children in grades 6 to 8 were asked to indicate their interest in 80 annotated book titles. The titles making the strongest appeal to the boys were concerned with violent adventure, travel, exploration, stories about boys, and biographies of men. For girls, the most popular subjects were love and romance, mild adventure stories with a child hero, descriptions of specifically feminine activities, and biographies of women. Sports on the whole interested boys more than girls, although the sex difference depended in part upon the specific sport. Girls chose stories about boys somewhat more often than boys chose stories about girls. Such reading interests reflect differences in the relative maturity of boys and girls, as well as more enduring sex differences in general interests. In a similar

survey conducted on Swiss school children (4), the principal interest of boys was the adventure story, that of girls the family story and biography. These sex differences in reading interests have been closely corroborated by extensive studies of the *movie* (62) and *radio* (19) preferences of boys and girls.

The *vocational choices* reported by children and adolescents reflect a similar dichotomy of interests between boys and girls. In one questionnaire survey (61), high school students were asked about their vocational preferences in terms of seven general job characteristics, rather than in terms of the traditional job classifications. Significant sex differences were found in each of the seven categories, the critical ratios ranging from 4.09 to 12.13 (cf. 86, p. 967). A greater percentage of girls expressed a preference for work entailing little responsibility, conducted indoors, and dealing with people rather than things. The boys more often chose work which involved: calmness rather than enthusiasm; risk or discomfort, but compensated by higher pay; planning *versus* carrying out another's plans; and directing *versus* following. It should be noted, of course, that such vocational preferences may merely reflect the student's realization of sex differences in vocational *opportunities*, rather than being a genuine expression of personal interest. In a comparison of the *"areas of life concern"* which high school students ranked highest for discussion and reading in school, certain sex differences were found which increased in late adolescence (82). The boys gave the highest ranks to discussions of physical health, safety, and money, and showed a more openly expressed interest in sex. The girls were more concerned about personal attractiveness, personal philosophy, planning the daily schedule, mental health, manners, personal qualities, and home and family relationships.

Investigations conducted on *adult groups* by a variety of techniques reveal similar sex differences in interests and attitudes. A number of investigators have analyzed the *conversations* of men and women by a method which may be unceremoniously described as "eavesdropping." Observers systematically recorded the topics of conversation overheard in New York's theatre district (63), on a midwestern college campus, in churches, hotel lobbies, streetcars, and other public places (47). A similar survey was made on two busy London streets (46). In another survey, three observers tallied the topics of conversation overheard during the ten-minute intermission at nineteen

concerts, over a six-month period (15). The results of this study, based on 601 samples of conversation among adults, were typical of those found by all the other investigators. Although the locale does to a certain extent determine the topics of conversation, the principal sex differences are quite consistent. Money, business affairs, and sports are more common in conversations between men; other women and clothes are more common in conversations between women. Moreover, women converse to a significantly greater degree than men about people. The conversations of mixed groups tend to be dominated by topics either of equal interest to both sexes or of little interest to either.

In a typical survey (12) on *newspaper reading*, the investigators found much in common between the reading interests of men and women. In reference to such differences as they did find, however, the authors concluded that women "tend to slight the things about news that matter in the social sense, and are most interested in the commonplace, ephemeral and human-interest sides of life." Sex differences in interest for different types of activity are indicated by such *interest tests* as the Kuder Preference Record (91, 92), designed especially as measures of vocational preferences. On the average, males show stronger preferences for mechanical, persuasive, computational, and scientific work. Female averages indicate greater interest in the literary, musical, artistic, social service, and clerical areas. Similar differences between men and women in general have been found on the Strong Vocational Interest Blank (81). On the other hand, groups of men and women engaged in the *same occupations* showed very similar interest patterns (77). Thus women physicians or life insurance saleswomen resembled men physicians or life insurance salesmen much more closely in their interests than they did housewives.

Significant sex differences have been obtained on the Allport-Vernon Study of Values (14). A psychograph constructed from the average scores of 1163 men and 1592 women is given in Figure 91. Women's responses rate highest in the aesthetic, social, and religious values. This suggests that the immediate enjoyment of artistic experiences, a concern for the welfare of other people, and an emphasis upon spiritual values may be relatively important in the life goals of women. The men's psychograph shows peaks in the theoretical, eco-

nomic, and political values. Such a profile indicates an interest in abstract knowledge and understanding, a drive for practical success, and a desire for prestige and power over others. These sex differences are not, however, very large, and the overlapping is con-

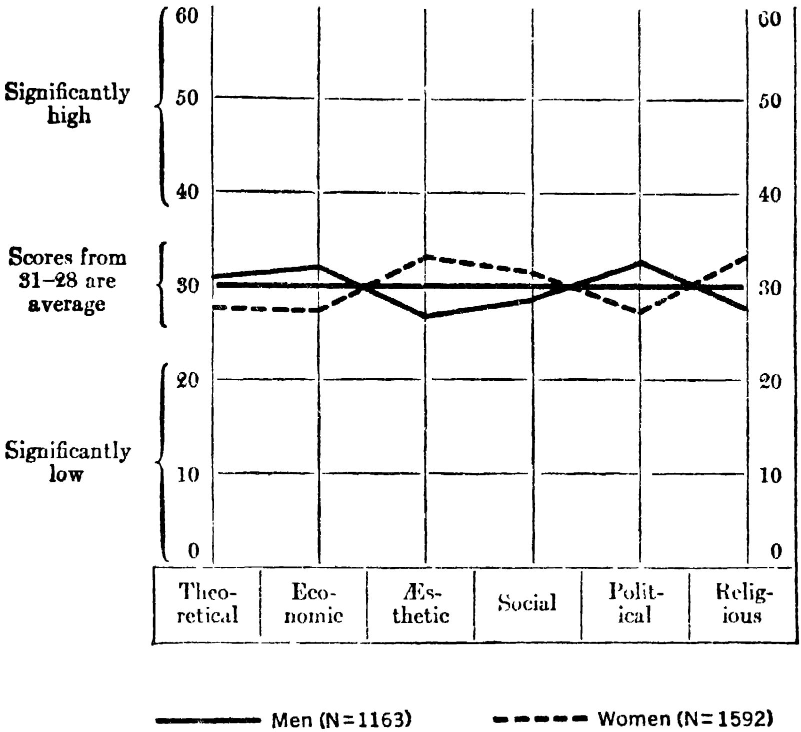


Fig. 91. Composite Profiles of Adult Men and Women on the Allport-Vernon Study of Values. (Data from Cantril and Allport, 14, p. 260.)

siderable. Far larger differences on this test have been found between different occupational groups of the same sex, than between men and women in general. Averages as low as 21 and as high as 49 have been obtained by men in different occupational groups, while the averages of men or of women as a group do not drop below 27 nor rise above 33 on any one value score. As in the case of the Strong Vocational Interest Test, occupation seems to introduce more of a difference in these scores than does sex.

SOCIAL AND EMOTIONAL CHARACTERISTICS

Social Adjustment. Certain consistent sex differences have been reported in the adjustment to social mores and restrictions, as well as in other aspects of personality commonly designated as character traits. In an extensive series of tests by Hartshorne, May, and Shuttleworth (33) on approximately 850 elementary school children in three cities, significant differences in favor of the girls were found in *moral knowledge and social attitudes*. Several tests of each of these aspects of character development were employed. In order to keep as close as possible to the children's own opinion, the tests were worked up in the form of ballots and the children were asked to "vote" on each item. In the so-called duties test, for example, several propositions were given with the request that the subject indicate whether it is his duty to do these things, by underlining *Yes*, *No*, or *S* (sometimes yes and sometimes no). Some of the items in this test were as follows (33, pp. 46-47):

1. To help a slow or dull child with his lessons	Yes	S	No
2. To call your teacher's attention to the fact if you received a higher grade than you deserved	Yes	S	No
3. To smile when things go wrong	Yes	S	No
4. To report another pupil if you see him cheating	Yes	S	No

In total scores on both the moral judgment and social attitudes tests, the differences in favor of the girls were 4.31 times as large as their standard errors and can therefore be regarded as highly significant. The investigators concluded that: "It appears on the surface at least that girls are more sensitive to both conventional and ideal social standards than boys" (33, p. 119).

Significant sex differences were also discovered in certain objective behavioral tests of character. In a series of investigations by Hartshorne, May, and Maller (31, 32), tests were devised in the following areas. "*deceit*," including cheating, lying, and stealing; "*service*," including cooperative and charitable behavior; and "*self-control*," including persistence and inhibition. Among the special advantages of these tests may be mentioned the fact that the subjects did not realize that they were being tested or that their actions could be detected. All observations, furthermore, were made in the course of ordinary everyday activities of the children, including school work, homework assignments, athletics, and party games. Data on deceit

were collected on some 10,865 elementary school pupils in several parts of the country. For the main studies on service and self-control, about 900 children were employed.

TABLE 41 *Sex Differences in Certain Character Traits*

(Adapted from Hartshorne, May, and Maller, 32, pp. 156, 380, 382)

Measure Employed	*Diff.*[a]/σ_{diff}
Total service score	1.9
Reputation for service	7.9
Total persistence score	1.7
Reputation for persistence	7.6
Total inhibition score	5.5
Reputation for inhibition	5.0

[a] All differences favor the girls.

No consistent sex difference in deceptive behavior was found. Analysis of separate tests showed that boys tended to be more honest in some situations, girls in others. In the studies of service and self-control, sex comparisons were made in both test scores and "reputation" among classmates and teachers. Summary data on sex differences in these tests are given in Table 41. It will be noted that all of these differences favor the girls. In service and persistence, however, the differences in total scores were not significant. The relative standing of the two sexes in these areas also varied from one test to another. Persistence scores depended largely upon the appeal of the specific subject matter for boys or girls. Other investigators have also obtained conflicting results with persistence tests. In inhibition, on the other hand, girls were significantly superior in total score and consistently superior on each separate test. The more successful adjustment of girls to the school situation may be partly the result of such a personality difference.

It is also interesting to note in this connection that in *reputation* the girls excel the boys markedly *in all traits*. This too may influence their school success. The discrepancy between reputation and performance is likewise of interest in relation to social pressure. It may be that with increasing age the cumulative force of social expectancy

becomes more effective and the discrepancy between behavior and traditional belief is lessened. With this would come an increasing differentiation between the sexes. Until similar behavior tests are made on adult subjects, these questions cannot be answered.

Another source of data on social adjustment is provided by statistics on *crime and delinquency.* Such records must, of course, be interpreted with considerable caution, since opportunities for crime are very different for the two sexes. Moreover, the differential treatment of the two sexes by the courts is clearly apparent. For most crimes, the available statistics probably underestimate the frequency of occurrence among women. The one exception is sex delinquency, which is judged with less leniency for women than for men. Whatever the reasons, however, the discrepancy between the crime records of the two sexes is tremendous. During a typical year, the men sent to federal and state prisons and reformatories outnumbered the women in the ratio of nearly 25:1 (cf. 73, p. 245). A similar ratio was found between male and female convictions in New York State within a one-year period. But when the number of arrests was considered, the ratio dropped to 19:1 (73, p. 248). The latter finding illustrates the differential treatment of men and women by the courts. Statistics on juvenile delinquency vary widely from one report to another, owing to such factors as the criterion of delinquency and differences in local conditions and practices. All agree, however, in showing a much greater proportion of delinquent boys than girls (86).

A similar excess of boys is found among the children referred to child guidance clinics as *behavior problems* (1). Additional data are furnished by a number of extensive school surveys in which teachers were asked to supply information about the problem children in their classes. The results of all these surveys are in close agreement (86). In one investigation covering ten cities, the ratio of boys to girls in the problem group was 4:1 (97). Among the undesirable types of behavior reported much more frequently for boys than for girls are: truancy, destruction of property, stealing, profanity, disobedience, defiance, cruelty, bullying, and rudeness (96). Moreover, a larger number of undesirable behavior manifestations per child are reported for boys than for girls (96). To what extent these sex differences may be a reflection of teachers' attitudes toward boys and girls, and to what extent they represent real behavior differences, is difficult to determine. That the differences are at least partly the result of a

"sex halo" in teachers' ratings is suggested by a number of investigations of such ratings (cf. 86, pp. 987-989).

Emotional Adjustment. The greater frequency of behavior problems among boys is probably related to a more general sex difference in *aggressive and dominant behavior*. The origins of this sex difference are probably partly cultural and partly biological. It will be recalled, for example, that a similar sex difference in aggressive and pugnacious behavior has been observed in many species of animals. The greater size and muscular strength of the male is undoubtedly one contributing factor, and the male sex hormone is another. The part played by the latter is demonstrated by the conspicuous changes in aggressive behavior following gonadal transplants in animals.

Whatever its origins, this sex difference is a particularly persistent one in our culture, having been observed from early childhood to adulthood. Studies on nursery school groups have repeatedly demonstrated that boys display anger and aggression more often than girls. In one investigation (34), for example, independent ratings by three teachers were obtained for each of 579 nursery school children. The results indicated that boys more often grab toys, attack others, rush into danger, refuse to comply, ignore requests, and laugh, squeal, and jump around excessively. Girls, on the other hand, more frequently exhibit withdrawing and introverted behavior, such as avoiding play, staying near an adult, seeking praise, and giving in too easily. A number of these differences were of questionable statistical significance, the critical ratios being under 2.58. At the same time, most of the observed sex differences were revealed as clearly among 2-year-olds as among 4-year-olds, a fact which led the author to minimize the role of social pressure in the greater aggression of boys. Direct observations of preschool children in standardized experimental situations have likewise shown a greater frequency of aggressive responses among boys (64). Quarrels with other children in kindergarten and elementary school are also more common among boys than girls (cf. 86).

The administration of personality tests of the questionnaire type has indicated similar sex differences in aggression or dominance among students and unselected adults. On the Bernreuter Personality Inventory, for example, the average male score in dominance has been found to be significantly higher than the female average in high school, college, and older adult groups. This test, consisting of 125

questions, can be scored with six different keys for as many different traits.[10] The average scores obtained by male and female college groups, together with the critical ratios [11] of the differences between them, are given in Table 42. The number of cases in these groups varied from 144 to 658. It will be noted that in the dominance scale the critical ratio of the difference in favor of males is 3.77.

TABLE 42 *Sex Differences among College Students on the Bernreuter Personality Inventory*

(Data adapted from Bernreuter, 7)

Scale	Male Average	Female Average	$Diff./\sigma_{diff.}$	Direction of Difference
B_1N: Neuroticism	57.3	−42.8	3.15	Women more neurotic
B_2S: Self-sufficiency	27.0	6.8	5.89	Men more self-sufficient
B_3I: Introversion	−25.6	−14.7	3.50	Women more introverted
B_4D: Dominance	45.9	30.6	3.77	Men more dominant
F_1C: Confidence	−51.5	8.7	9.62	Men more self-confident
F_2S: Sociability	−25.9	−31.1	0.88	Women more gregarious and socially dependent

Another area of behavior showing large and persistent sex differences is that of *social orientation*. Some evidence has already been presented—in our discussion of interests, preferences, and attitudes—which indicates a much stronger interest in people among women than among men. This sex difference also appears early in life and continues into old age. One possible factor in the greater social interest and social orientation of girls may be their earlier language development. Their more rapid mastery of speech would certainly give girls an advantage in communicating with other children as well as with adults, and would thus encourage activities of a social nature. Of prime importance, however, are the subtle social pressures which probably begin to operate much earlier than is generally realized. Traditional sex roles and sex stereotypes are almost certain to be reflected in the attitudes of parents and others toward the child almost from the time of his birth.

[10] The fact that some of these traits have been shown by factor analysis to be correlated with each other, and that the last two represent common factors identified through the first four, introduces unnecessary duplication in the scores, but does not invalidate the sex comparisons made. The six scales simply represent six categories into which the responses can be grouped.

[11] Computed by the writers from the norms published by Bernreuter (7).

Throughout childhood, sex differences in sociality have been noted in a wide variety of situations (42, 86). In the play activities of nursery school children, boys show more concern with things, girls with personal relationships. Similarly, girls manifest more responsibility and "motherly behavior" toward other children than do boys. At all ages, girls engage more often in "social" games involving other children; they read more books about people and more frequently express interest in occupations dealing with people. The girls' greater concern with questions of appearance and manners is indirectly an indication of more interest in what others will think of them. Parents' tabulations of the questions which children asked in their presence showed a significantly greater proportion of questions about social relations asked by girls. Nicknames of an affectionate form are more common among girls, those based on physical peculiarities more common among boys. Girls are more frequently angered by situations affecting their social prestige, and also experience more jealousy. Even studies of children's dreams have shown that girls more often than boys dream about people of various sorts, as well as about their own family and home.

In interviews with 666 children between the ages of 5 and 12, responses involving persons, family, and social relations occurred with consistently greater frequency among the girls than among the boys (41). Among the children's "first wishes," for example, the proportion dealing with siblings, companions, or friends was 12% for girls and 3% for boys. In descriptions of "the best thing that ever happened" to them, 14.0% of the girls and 8.3% of the boys mentioned parental contacts and other personal relationships. Corresponding percentages for "the worst thing that ever happened" to them were 15.6 and 8.0. At the opposite extreme of age, it is interesting to note that among persons between 70 and 90 years old, sociability showed a high positive correlation with happiness in females, but an insignificant correlation in males (42).

An important distinction has been made by Johnson and Terman (42) in their discussion of sex differences in social orientation. Data collected on 3000 college students suggested that women do not actually behave more socially, although they desire more strongly than men to be social. The overt expression of social interests in the female is more often inhibited by timidity and lack of self-confidence, as well as by more specific culturally imposed restrictions. The previ-

ously discussed sex difference in aggression probably affects social participation. As for self-confidence, a number of studies have indicated that women rate lower than men in this regard. For example, the largest and most highly significant sex difference reported in Table 42 was found in the confidence scale of the Bernreuter Personality Inventory. The critical ratio of this difference is 9.62. A similar difference in favor of males, with a critical ratio of 5.89, was obtained in the self-sufficiency scale.[12] On the other hand, the lack of significant sex difference in the "sociability" scale of the same inventory may be due to the inclusion of two aspects of sociability showing opposite sex differences, viz., social interest and social participation.

The distinction between social interest and social participation may also help to explain some of the ambiguities in the results on sex differences in introversion. We have already seen (Ch. 15) that common tests of introversion measure more than one "unitary trait." The factor analyses of Mosier and the Guilfords suggested a convenient tripartite division into "thinking," "social," and "public" introversion. For a comparison of the sexes, a different sort of distinction may be necessary, possibly between social orientation or interest and social contacts or participation. Turning again to Table 42, we find that women appear significantly more introverted on the average than men. Surveys with the Guilford and Martin inventories on high school students and rural adults showed no significant sex difference in "thinking" introversion (30). "Social" introversion gave inconsistent sex differences in the high school and rural groups.

The first clear-cut evidence that sex differences may cut across the traditional introversion-extroversion category is to be found in a study by Heidbreder (36). A list of 54 "introvert traits" was carefully compiled so as to include the behavior most frequently designated as characteristically introverted. Self-ratings and ratings by two associates on each of these traits were obtained for 100 college men and 100 college women. No significant sex difference was found in total introversion scores, the averages being 11.41 and 11.12 for the men and women, respectively. But the introvert characteristics reported most frequently by the men differed from those reported most fre-

[12] This should not be regarded as independent corroboration, since the two scales overlap considerably. It does suggest, however, that it is the self-confidence items which account for the sex difference, rather than some other unidentified aspect of behavior which might have been involved in the test score.

quently by the women. A few examples will serve to illustrate this difference.

Typical "masculine" symptoms of introversion:

Outspoken
Works things out on own hook; hesitates to accept help
Keeps in background on social occasions
Conservative and painstaking in dress
Introspective

Typical "feminine" symptoms of introversion:

Shrinks when facing a crisis
Works by fits and starts
Has ups and downs in mood without apparent cause
Feels hurt readily
Hesitates in making decisions on ordinary matters

Another point should be considered in evaluating sex differences in introversion among college students. There is some evidence suggesting that introversion scores tend to be positively related to academic success among boys, but negatively related to academic success among girls (cf. 42). Those students who have been sufficiently successful in their school work to reach college are thus likely to show a smaller sex difference in introversion than is found among unselected adults. Moreover, the environment of male and female college students is undoubtedly more similar than that of unselected adult men and women. In summary, what the results on introversion show is that neither men nor women can be said to be more introverted. Any sex differences reported in measures of introversion must be interpreted with reference to the particular population studied, as well as the *specific behavior manifestations* which were sampled. Existing sex differences can evidently be better described in terms of social orientation than in terms of the traditional introversion-extroversion category.

A third major personality area in which large sex differences have been reported is that of *emotional instability* or *neuroticism*. Observations of preschool and elementary school children have revealed a somewhat greater frequency of "nervous habits," such as nail-biting and thumb-sucking, among girls than among boys (cf. 86). It should be noted, of course, that "nervous habits" represent a rather

arbitrary behavior category. The greater frequency of "behavior problems" among boys may balance the greater frequency of "nervous habits" among girls. The total degree of instability might thus be no different in the two sexes at these age levels. Girls may simply resort to milder and less violent ways of expressing displeasure and malad-

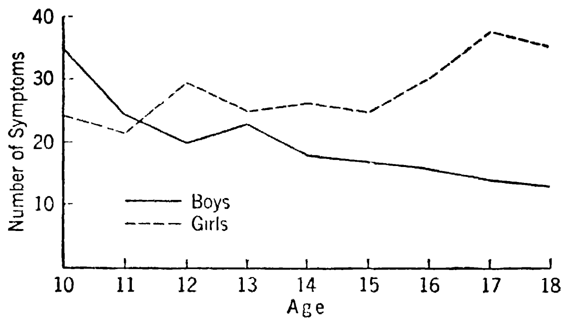


Fig. 92. Median Number of Symptoms Reported by Boys and Girls on the Woodworth-Mathews Test of Emotional Instability. (From Mathews, 56, p. 21.)

justment than boys, because of differences in socially imposed restrictions.

Fear responses have also been found to occur more commonly among girls, as determined by laboratory investigations, teachers' reports, and interviews with children (34, 41, 86). A culturally determined sex difference in the *admission* of fear may exaggerate these differences somewhat. Girls also tend to report more worries, as well as emotional responses of greater intensity and to a wider variety of stimuli than do boys (84, 86).

On neurotic inventories, clear-cut sex differences in emotional instability do not appear until the adolescent years. This finding was corroborated in a number of investigations with adaptations of the Woodworth Personal Data Sheet, specially designed for use with children and adolescents (32, 56, 85, 86). An interesting illustration of age changes in this respect is shown in Figure 92, based upon the scores of 575 boys and 558 girls between the ages of 9 and 19. At age 10, the boys reported a larger median number of neurotic symptoms than the girls on the Woodworth-Mathews Test. With increasing age, the median number of symptoms tends to rise among the

girls, but drops among the boys. Beyond age 14, the sex difference is statistically significant and consistently in favor of boys. The increasing differentiation of social pressures with age is one hypothesis that is obviously suggested by such data.

Among adult groups, sex differences on neurotic inventories are large and consistent. Again referring to Table 42, we find a critical ratio of 3.5 for the sex difference on the neuroticism scale. Similar evidence for greater female emotionality was found in the previously cited Guilford and Martin study on high school students and rural adults (30). That such differences are real and not limited to questionnaire replies is suggested by a study of a college group (17). Students who had taken a personality inventory were subsequently interviewed by two experienced counselors. The excess of maladjustment among the women, as revealed by the interviews, was even greater than that indicated by the test scores.

In their analysis of the relevant literature, Johnson and Terman (42) tend to emphasize constitutional rather than cultural factors as a basis for the greater emotional imbalance of the female. Among the types of evidence which they cite in support of such a conclusion are: (1) the early age at which sex differences in nervous habits appear; (2) the persistence of sex differences in neuroticism today, despite the trend toward equalization of social pressures; (3) the fact that sex differences in emotionality are as great or greater among institutionalized blind, deaf, or orphaned children, despite the relative uniformity of the institutional environments; and (4) the fact that peaks of "nervous" behavior often coincide with such physiological changes as puberty and the menopause.

All these lines of evidence must be interpreted with caution. Puberty and the menopause are periods of acute social crises in our society, as well as periods of physiological upsets. Institutional environments are far from uniform for boys and girls.[13] In fact, there is no reason to suppose that sex stereotypes are any different among institutional personnel than among any other members of our culture. The persistence of sex differences in neuroticism in contemporary society is not surprising. The greater equalization of education and the sporadic admission of women to certain predominantly "masculine" occupations, without the removal of other sources of frustration and discrimination, may increase rather than decrease conflict

[13] Cf. definition of environment in Chapter 4.

and maladjustment. As for the excess of nervous habits among female children, it has already been pointed out that nervous habits may be an insufficient index of emotionality. The evidence does not conclusively show that the female is the more "emotional" sex in childhood, even if we were to grant that the environments of boys and girls were equated.

A "MASCULINITY-FEMININITY INDEX" OF PERSONALITY

An approach to sex differences which has been used increasingly in recent years is the comparison of men and women in those responses which have proved to be most characteristic of each sex in our contemporary culture. Test items are chosen on the basis of their ability to discriminate between the responses of the sexes. Thus if 30% of the men and 29% of the women were to report that they like modern art, the item would be discarded because it does not differentiate between the sexes. Similarly, if 76% of the men and 79% of the women dislike walking in the rain, this item is also eliminated. Only those items marked by a significantly different proportion of men and women are retained. The resulting test provides an index of "masculinity-femininity" in the sense that it reflects the characteristic male and female responses in our culture. This approach is illustrated by the masculinity-femininity scores on such tests as the Strong Vocational Interest Blank, the Minnesota Multiphasic Personality Inventory, and the Interest-Attitude Analysis prepared by Terman and Miles.

It should be noted that such tests are deliberately designed so as to exaggerate sex differences, in the same way that such intelligence tests as the Stanford-Binet are designed to exclude or minimize sex differences. The behavior of men and women undoubtedly shows many similarities. These tests, however, concentrate on the differences, since it is their purpose to measure the differences between men and women as fully as possible. For any person taking such a test, the M-F (masculinity-femininity) index indicates the degree to which his responses agree with those most characteristic of men or of women in our culture. It is customary to designate either the masculine or the feminine end of the scale arbitrarily as + or −, for purposes of quantification.

The most extensive investigation of characteristic sex differences in personality is that conducted by Terman and Miles (84). After an exhaustive survey of the literature and prolonged research, items were chosen which revealed the most pronounced differences between representative samplings of men and women in our society. Data were gathered on many hundreds of persons, including elementary school, high school, college, and graduate students; unselected adults; members of several occupations; and specially selected groups such as athletes, juvenile delinquents, and adult homosexuals. The Interest-Attitude Analysis, constructed as a result of this research, consists of seven parts: Word Association, Inkblot Association, Information, Emotional and Ethical Attitudes, Interests, Opinions, and Introvertive Response.

This scale proved very successful in differentiating between the responses of male and female groups. Significant sex differences in total score were obtained at all age levels, from teen-agers to octogenarians. The critical ratios of these differences ranged from 7.2 to 39.9. Overlapping of male and female distributions was also relatively slight. The test thus achieved its purpose of selecting those behavior characteristics which differentiate most clearly between the sexes.

An intensive analysis of the male and female responses on each part of the test brought to light those aspects of the personalities of the two sexes which are most clearly differentiated in our culture. Terman and Miles summarize these differences as follows:

> From whatever angle we have examined them the males included in the standardization groups evinced a distinctive interest in exploit and adventure, in outdoor and physically strenuous occupations, in machinery and tools, in science, physical phenomena, and inventions; and, from rather occasional evidence, in business and commerce. On the other hand, the females of our groups have evinced a distinctive interest in domestic affairs and in æsthetic objects and occupations; they have distinctly preferred more sedentary and indoor occupations, and occupations more directly ministrative, particularly to the young, the helpless, the distressed. Supporting and supplementing these are the more subjective differences—those in emotional disposition and direction. The males directly or indirectly manifest the greater self-assertion and aggressiveness; they express more hardihood and fearlessness, and more roughness of manners, language, and sentiments. The females express themselves as more compassionate and sympathetic, more timid, more fastidious, and æsthetically sensitive, more emotional in general (or at least more expressive of the

four emotions considered), severer moralists, yet admit in themselves weaknesses in emotional control and (less noticeably) in physique (84, pp. 447–448).

In regard to the *origin* of such sex differences in personality, there are several lines of evidence which suggest the greater role of cultural than biological influences. One source of relevant data is to be found in some of the group profile comparisons reported by Terman and Miles (84, pp. 570–579). These profiles, showing the sub-test averages of various male and female groups, strongly suggest the *specificity* of differences in masculinity-femininity. Groups with the same mean total score may achieve such a score in very different ways. For example, among the most "masculine" groups in terms of M-F index are high school boys and engineers. Both obtained identical mean total scores, but the high masculinity of the high school boys resulted largely from their interests and information, while that of the engineers was primarily due to their emotional and ethical attitudes. On the latter test, the high school boys were actually more feminine than the general male population. Similarly, groups of delinquent girls and of women artists received mean total scores which coincided with the norm for the general female population. The delinquent girls, however, achieved this result by a very "feminine" performance on the test of emotional and ethical attitudes, and a very "masculine" performance on the interest test. The women artists, on the other hand, were significantly more "feminine" than the general female norms in interests, but significantly more "masculine" in information. Men artists, with a similar deviation in the "feminine" direction in interests, as well as in the word association test, received a much more "feminine" total score than the male norm, although their scores on the remaining sub-tests were at the general male average.

Correlations between M-F scores and *physical characteristics* have been generally low and insignificant (84, Ch. V; 27). Such correlations as have been found are probably the result of the social effects of certain conspicuous physical characteristics, rather than the result of underlying biological factors. For example, a slight tendency has been found for taller men (84) and for men with deeper voices (27) to obtain a more masculine M-F index. Such a correlation may simply reflect the influence of social stereotypes upon the development of the

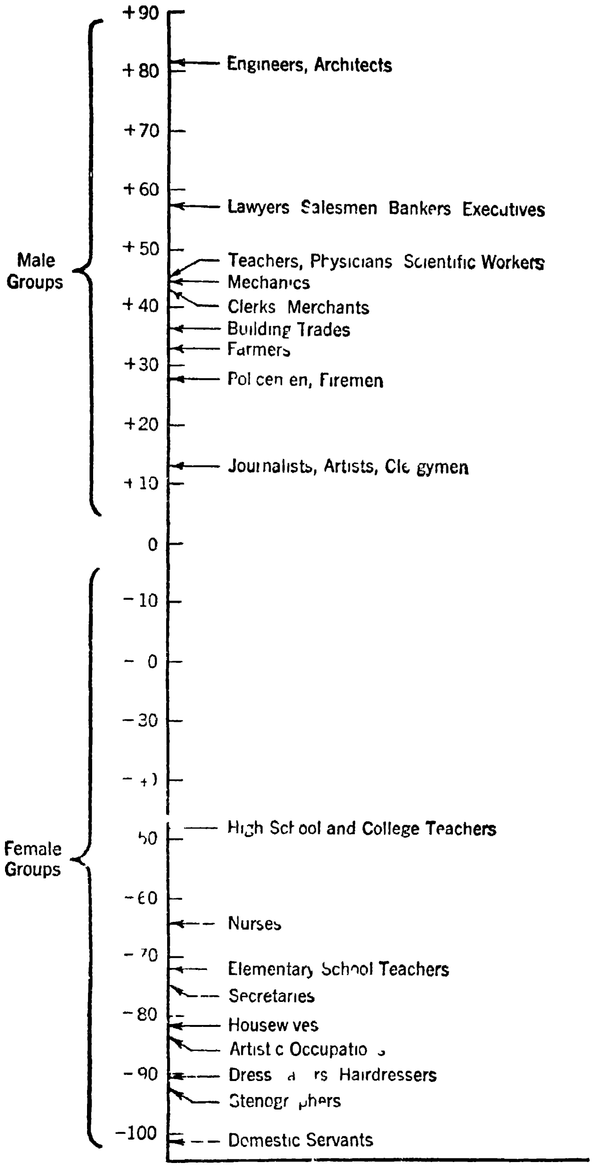


Fig. 93. Mean M-F Index of Men and Women in Various Occupations. The positive end of the scale indicates the more masculine scores, the negative end the more feminine scores (Data from Terman and Miles, 84, pp 159, 179)

individual's personality. Studies of male homosexuals (84, Chs. XI–XIII) have also indicated that experiential rather than structural factors were primarily responsible for the development of homosexual behavior. Especially important were early home environment and parental attitude toward the individual.

In the general population, the M-F index has been found to be significantly associated with *education* and *occupation* (84). Illustrative data on occupational groups are shown in Figure 93. It was also found that highly intelligent and well-educated women tend to score more "masculine" than their sex norms. For example, women listed in *Who's Who,* as well as those holding an M.D. or a Ph.D. degree, average more "masculine" in total score than any of the occupational groups shown in Figure 93. Similarly, men who have cultivated avocational interests of an artistic or cultural nature tend to obtain more "feminine" scores. Thus the equalizing influence of specific training or experience seems to bring about a convergence of the temperamental qualities of the two sexes.

The M-F index seems also to depend upon the *domestic milieu* in which the individual was brought up. Such factors as the death of one parent, excessive or exclusive association with one or the other parent, and predominance of brothers or of sisters among the siblings are much more closely related to M-F score than are physical traits (84). Moreover, there is some evidence to suggest that deviation toward the norm of the opposite sex in both men and women is associated with unpleasant and undesirable childhood experiences, broken homes, and parental maladjustment (23, 78). A pleasant, happy childhood, on the other hand, encourages the individual to accept the appropriate male or female model of behavior presented by his culture.

To summarize, in personality as in intelligence, we cannot speak of inferiority and superiority, but only of *specific differences* between the sexes. These differences are largely the result of *cultural and other experiential factors,* although certain physical sex differences may influence behavior development, either directly or through their social effects. Lastly, the overlapping in all psychological characteristics is such that we need to consider men and women as *individuals,* rather than in terms of group stereotypes. These three points will prove to be useful "rules" to observe in understanding other group differences to be considered in the chapters which follow.

CHAPTER

20

Racial Comparisons: Problems of Grouping

THE COMPARATIVE EVALUATION of the races of man has long been a subject of keen interest and lively controversy. It is an interesting commentary upon human thought that nearly all theories of racial inequality proclaim the superiority of the particular race of their respective exponents.[1] Thus Aristotle (cf. 46, pp. 318-320) endeavored to demonstrate that the intellectual leadership of the Greeks must of necessity follow from their favorable geographical location. He argued that the peoples inhabiting the colder regions of northern Europe, although outstanding for bravery and physical prowess, were intellectually incapable of a high degree of political organization or leadership. Similarly, the Asiatics, although intellectually keen and inventive, lacked spirit. The Greeks alone, being geographically intermediate, were endowed with the proper balance of these traits and were thus by nature fitted to rule the earth. Similar claims have been made for such groups as the Arabians, the Romans, the French, the Anglo-Saxon, the "white" race as distinguished from those having a different skin pigmentation, the Nordics, the Alpines, the Mediterraneans, and various others.

Outstanding among such theories, because of its widespread popularization, is that proposed by de Gobineau (17) in the nineteenth century and subsequently expanded by Chamberlain (12). This doctrine had numerous followers who reformulated it and developed it along various lines. Its basic contention, however, is the superiority of the Nordic or "Aryan" race, a loosely and ambiguously defined group whose descendants are now supposed to inhabit for the most part the countries of northern Europe. The array of evidence cited in support of this theory is incomplete and one-sided at its best and

[1] For a readable historical survey of theories of "racial superiority," cf. Benedict (2).

fantastic and mythical at its worst. The concepts involved in such a theory will be critically examined in the course of the subsequent discussion.

Within our own generation, race problems have flared up with violent intensity and shocking effects. Outworn and forgotten theories have been revived in an attempt to rationalize political actions and policies. The Nazi racial doctrines during World War II represent a recent illustration of such a perversion of anthropological material. Probably the first writer to use the data of physical anthropology for nationalistic propaganda was de Quatrefages who, during the war of 1870, referred to the Germans as "Huns" (cf. 13). The latter term was revived as a derogatory epithet during World War I.

Under the stress of emotional appeal, it is especially difficult to carry on unbiased and objective analysis of facts. It is one of the earmarks of prejudice to draw logically unwarranted inferences from the data at hand. A typical testing technique for the measurement of the prejudice-fairmindedness variable, for example, is based upon just such behavior (cf., e.g., 59). The subject is given certain facts bearing upon controversial issues, with the instructions to check any of the proposed conclusions which seem to him to follow *directly* from the given data, regardless of their truth or falsity in general. The individual who is biased or who responds emotionally to any of the issues involved will ignore the limitations of the facts actually presented and will generalize far beyond them. The procedure in this test presents a close parallel to what probably occurs all too often in the interpretation of data on such emotionally toned issues as race differences.

Under such conditions it is especially important to recognize and to bear clearly in mind the possible vitiating factors and sources of error in the data. As in all group comparisons, studies on race differences must take into account *selective factors* and adequacy of sampling, *overlapping* of distributions, *reliability* of an obtained difference, *inaccuracy or ambiguity of the measuring instrument*, and other similar factors which have already been discussed and illustrated (cf. Ch. 18). It is probably not an exaggeration to state that failure to consider such factors has invalidated the large majority of investigations which purport to have established a racial difference in one or another ability or personality characteristic.

Racial comparisons are an extremely difficult problem of differ-

ential psychology. In addition to the above-mentioned sources of error which they share with all group comparisons, studies on race differences are handicapped by special difficulties inherent in every phase of the problem. Thus it has proved a difficult matter in such studies to decide *whom* to measure, *what* to measure, and *how* to measure it. These difficulties will be analyzed in the present and subsequent chapter. The first of these two chapters will be concerned with questions of *whom* to measure, or the selection and classification of subjects in racial studies. In the second chapter will be discussed some of the major problems which arise in the efforts to *measure* and *compare* widely diverse groups. A third chapter will be concerned with the relative contribution of racial and cultural factors to the development of existing group differences in behavior. The special experimental designs which have been employed for this purpose will be considered and illustrated with typical investigations.

The data of investigations on race differences have been grouped about these methodological questions. No general summary of findings and no "intellectual hierarchy" of racial IQ's are presented because, although apparently useful as mnemonic devices, such tabulations would be of dubious value. Isolated facts are particularly misleading in racial comparisons and should at all times be evaluated in terms of the conditions under which they were collected. Conclusions on race differences will therefore be drawn only in the light of a critical analysis of the entire problem and will not be divorced from their limiting conditions.

No attempt has been made, furthermore, to survey the vast array of investigations on psychological differences among racial groups. For summaries and more extensive discussions of this problem, the reader is referred to such sources as Garth (25, 26), Klineberg (36), Mann (44), and others (56). Special surveys of psychological investigations dealing with the American Negro have been prepared by Klineberg *et al.* (37) and Canady (8). For an orientation to the general problem of race, books by Boas (5, 6), Kroeber (41), and Dunn and Dobzhansky (20) will prove helpful. To obtain a well-rounded picture of specific groups, the reader may consult the intensive field studies conducted by anthropologists, sociologists, and psychologists on certain groups, such as C. du Bois' study of the Alorese in the South Pacific (19), or the investigation of Kluckhohn and Leighton (38, 39) on the Navajo Indians. Such studies report

psychological test results as well as a detailed analysis of the cultural background of the groups. The same approach is illustrated by the extensive series of investigations on Negro youth conducted by the American Youth Commission of the American Council on Education (1, 16, 23, 34, 47, 54, 57). Although primarily sociological in methodology, the latter studies contain a wealth of psychological material. Data were gathered on several thousand Negro adolescents in many parts of the United States. Community studies, intensive interviews, case studies, and psychological testing were among the methods employed in different parts of the survey.

WHAT IS A RACE?

Tradition, prejudice, and the snap judgments of everyday observation have contributed to the development of a concept of race as a clearly differentiated and easily identifiable group, possessing distinctive physical, mental, and temperamental characteristics. The observations of biologists, anthropologists, and psychologists, however, fail to support such a view.[2] The classification into racial groups is essentially a biological one and corresponds to such divisions as breed, stock, and strain in infrahuman organisms. In its simplest terms, any definition of race implies a certain *community of physical characteristics based primarily upon a common heredity.*

The task of race classification is far more complex than would appear from the glibness with which individuals are commonly assigned to one group or another. The fivefold classification of races, formerly memorized by every school child, is of historical interest only. This system can be traced to Linnæus (43), the great classifier, who recognized four races of men—*Europæus albus* (white), *Americanus rubescens* (red), *Asiaticus fuscus* (yellow), and *Africanus niger* (black). A fifth group, the brown race, was subsequently added by Blumenbach (3), who also altered the terminology, proposing the now familiar classification into Caucasian, Mongolian, American, Ethiopian, and Malayan. This classification is crude and superficial, as will shortly become apparent.

The essential problem in the classification of racial groups consists in the identification of inheritable physical characteristics which differ

[2] For a very readable account of many of the difficulties of race classification, see Huxley and Haddon (33).

clearly from one group to another and which may thus serve as *criteria* of race. A wide variety of such criteria have been proposed and applied (cf., e.g., 13, 36, 41). *Skin color*, although popularly employed as one of the most obvious means of racial identification, has proved to be one of the least satisfactory of the possible criteria. It is a well-established fact that the same pigments are present in all human skins and that different skin colors result from varying relative amounts of each pigment. For this reason, there is found a complete series of transition shades, making exact classification very difficult. Such a classification is also rendered somewhat unstable by the fact that environmental conditions, such as exposure to the sun's rays, have a marked effect upon skin color.

Pigmentation of the eyes has proved to be a somewhat more promising index, in so far as it is unquestionably a hereditary trait. In the same connection may be mentioned *hair color*. These traits, however, are also difficult to describe quantitatively because of continuous gradations. A further difficulty in the use of such criteria is their relatively narrow distribution, black hair and eyes being the universal rule outside of the Caucasian stock.

In addition to coloring, other characteristics of the hair have been employed as differentiating signs. The *texture of the hair* is generally regarded as a valuable aid in racial classification. For example, the straight, stiff hair of the American Indian is in sharp contrast to the woolly, tuft-like hair of the Hottentot. Fullness of the beard and *hirsuteness*, or amount and distribution of hair on the body as a whole, have also been employed in such classifications.

Racial groups have been differentiated on the basis of *gross bodily dimensions*, chief among which is stature. Group differences in this respect are, however, surprisingly small and consequently of doubtful value in racial identification. *Facial* and *cranial measurements* have been employed to somewhat better advantage. Among the former, the most common are nasal index, which expresses the relative length and breadth of nose, and various indices of prognathism, or the degree of protrusion of the jaws. Cranial capacity, or volume of the skull, yields rather ambiguous results because of its dependence on general body size and because of the wide variation *within* groups with consequent overlapping *between* groups. Cephalic index,[3] on

[3] Cephalic index $= \frac{100 \times \text{head width}}{\text{head length}}$. For a fuller description, see Ch. 12.

the other hand, has proved to be one of the most satisfactory criteria of classification and is now widely employed.

In view of the relative paucity of satisfactory anatomical criteria, attempts have been made to evolve physiological or biochemical schemas of classification. It has been suggested, for example, that races might be classified on the basis of *blood groups*, which have become familiar in connection with blood transfusions (cf. Ch. 16). These blood groupings refer to the agglutinative reactions of the red blood corpuscles, i.e., the tendency of such corpuscles to clump together when the blood of certain individuals is mixed with that of certain other individuals. In a few early studies, the relative incidence of A, B, AB, and O blood types in different racial stocks was used as a basis for racial classification, but the resulting groupings conflicted sharply with other criteria of race (31, 41). More promising results have been obtained by including the recently identified M, N, and Rh factors, as well as the various subgroups found for some of the factors. The racial classification suggested on this basis seems to agree fairly closely with the data on geographical distribution and common descent.

The *endocrine glands* have also played their part in race classification. Likenesses have been noted, for instance, between the physical and alleged psychological characteristics of certain racial groups on the one hand, and the characteristics associated with certain pathological glandular dysfunctions on the other (cf., e.g., 35). Thus a parallel has been drawn between the cretin and the African Pygmy. Pituitary enlargement has been attributed to the Hottentots and adrenal deficiency to the Negro. The "childlike" appearance of the Chinese has been ascribed to an overactive thymus. Such methods of classification are especially questionable for two reasons: they take a superficial and partial resemblance as their point of departure; and they reason from pathological conditions existing within a single group to the normal characteristics of entire groups.

Finally, mention should be made of the efforts to deal with race in terms of *constitutional type* (cf. Ch. 13). Kretschmer (40), for example, believed the ratio of leptosomes and pyknics to differ in various racial groups and offered this as a possible explanation for the psychological differences between such groups. Others, both prior and subsequent to Kretschmer, have attempted similar classifications. The reader should recall in this connection the difficulty of finding "pure types" and the absence of valid evidence for a correlation

between the physical characteristics of such types and any of their alleged psychological differentia.

EVALUATION OF THE CRITERIA OF RACE

In addition to the special deficiencies of individual methods of classification discussed in the preceding section, certain major difficulties are encountered in the application of all, or nearly all, criteria of race. In the first place, a wide *variability* exists within any one group in respect to any trait. Closely related to this is the marked *overlapping* between different groups in any of the criteria mentioned. Thus, although two groups may differ significantly in average height, individuals can readily be found in the "shorter group" who are taller than certain individuals in the "taller group." This obviously makes group delineations indistinct and relatively arbitrary.

A third difficulty is the *inconsistency* frequently found when more than one criterion is employed. An individual might have the coloring of a Nordic, the cephalic index of an Alpine, and the stature of a Mediterranean. Or very dark skin pigmentation and woolly hair might be found in association with Caucasian features. Such instances are frequent and cannot be dismissed as exceptions.

Finally, it should be noted that many of the alleged racial characteristics which were formerly believed to be stable and innate are being found to be susceptible to *environmental influences*. Even such apparently "hereditary" traits as body height, skull shape, and facial conformation have proved to be dependent in part upon environmental conditions in early childhood. This was illustrated in certain investigations by Boas (4, 7) on the American-born children of immigrants from several European countries. These children were compared with foreign-born children from the same countries, who were also living in America. Differences were found in stature, weight, and length and width of head.

The most striking demonstration of environmental influence, however, was furnished by an examination of the cephalic indices of two immigrant groups which differ markedly in head shape (4, 7). American-born and foreign-born boys were compared within an East European Jewish group and a Sicilian group, both living in New York City. The former are characteristically round-headed, having a high cephalic index; the latter are characteristically long-headed. As will be seen

in Table 43, residence in the new environment tends to make the Jewish group more long-headed and the Sicilian more round-headed, both groups converging toward the American norm It will also be noted that those boys born after a relatively long period of American residence of the mother tend to show a greater change than those born after a shorter residence period This was also found to be the case in the data on other immigrant groups.

TABLE 43 *Change in Cephalic Index of Two Immigrant Groups*

(After Boas, 4, p. 10)

Group	*N*	*Average Age*	*Average Cephalic Index*
Foreign born Sicilian boys	241	9.6	79.5
American born Sicilian boys			
Born less than 10 years after arrival of mother	375	10.0	80.0
Born 10 or more years after arrival of mother	127	9.5	81.8
Foreign born Jewish boys	179	9.1	84.6
American born Jewish boys			
Born less than 10 years after arrival of mother	257	9.2	82.4
Born 10 or more years after arrival of mother	290	9.2	82.3

That these physical changes were the result of changing environmental conditions rather than selective factors was clearly demonstrated A comparison of foreign born persons who had immigrated at different periods showed no significant differences in the traits under consideration The measurement of American-born and foreign-born children *of the same parents*, furthermore, revealed differences in the expected direction

The results of Boas have subsequently been corroborated by Guthe (29) who compared the cephalic indices of 1[illegible]7 Russian-born Jewish children and 127 American-born Jewish children in Boston The cephalic indices found by Hirsch (30) on American born children of South Italian, Russian-Jewish, and Swedish parentage were also in general agreement with the corresponding figures reported by Boas Similar changes in the shape of the head were found by Dornfeldt (18) through extensive measurements of migrating Jewish groups in Europe.

Investigations along such lines have also been conducted on Japanese groups migrating to Hawaii (50) and to the United States (53). Spier (53) obtained a series of anthropological measures on 320 American-born Japanese school children in Seattle, Washington, and its vicinity. The same measurements were repeated on 521 school children living in those sections of Japan from which most of the Seattle group was believed to have come. In general, the American-born children were larger, taller, more round-headed, and had wider faces than those born in Japan. Many of the individual comparisons of corresponding age and sex groups yielded statistically significant differences between the American-born and native subjects. As in the case of the European immigrants, the differences tended to become more marked the longer the mother had lived in this country.

A variety of factors have been proposed to account for the changes in physical type found in immigrant groups. Differences in bedding and cradling, as well as the gradual abandonment of swathing customs practiced in the mother country, have been cited as possible explanations of the changes in head shape. Nutrition and type of diet are doubtlessly important factors in all the physical changes noted. Alteration in the activities of the endocrine glands under the stress of adjusting to a new culture has also been suggested as a possible factor (cf. 30). Most of these explanations are, to be sure, highly speculative. Whatever the specific influence or influences at work, however, it is quite clear that they are of an environmental nature.

A TENTATIVE CLASSIFICATION OF RACIAL GROUPS

It is apparent that no one criterion of race can yield a satisfactory classification. Nor can clear-cut group distinctions be made with a combination of such criteria. It should be borne in mind that at best any racial classification is approximate. No sharp line of demarcation can be established between groups, nor can every individual be unequivocally assigned to one particular group. The classification which has been most widely used by anthropologists and psychologists is one based upon a *combination* of criteria, chief among which are cephalic index, hair quality, hairiness on the body, facial conformation, and bodily proportions. An outline of this classification (cf. 41, p. 132) is shown below:

I Caucasian
- Nordic
- Alpine
- Mediterranean
- Hindu

II Mongoloid
- Mongolian
- Malaysian
- American Indian

III Negroid
- Negro
- Melanesian
- Pygmy Black
- Bushman

IV Of doubtful classification
- Australoid
- Polynesian
- Ainu
- Veddoid (Indo-Australian)

Within the Caucasian or white race, four subdivisions are generally recognized. Three of these groups are the Nordic, Alpine, and Mediterranean classes into which the population of Europe is divided; the fourth consists of the Hindus. The Nordics are described as tall, blond, blue-eyed, fair-skinned, and dolichocephalic, or long-headed. They occupy a horizontal belt around the Baltic and North Seas, covering most of England, northern France, the Scandinavian peninsula, Holland, and northern Germany. The Alpines, located chiefly in central Europe, are of medium stature and intermediate coloring, but definitely brachycephalic, or broad-headed. In the Mediterranean group, we again find a pronounced dolichocephaly, accompanied by black or brown hair and eyes, relatively dark skin, and short stature. As its name implies, this group is found principally on the shores of the Mediterranean, comprising most of the population of Spain and Portugal, southern France, southern Italy, Greece, and certain parts of northern Africa. The Hindu, although darker skinned, bears a very close resemblance to the Mediterranean and is sometimes included within this group.

The Mongoloid race is characterized by straight hair, very little hair on the face and body, thin lips, and frequently the epicanthic fold which produces the appearance of "oblique" eyes. Short limbs are usually the rule in this group. Skin color may be yellow, brown, or

reddish. This race comprises the Oriental Mongolian, as well as the American Indian and the Malaysian. All three are believed to have evolved by differentiation of the same primary stock. Close and extensive observation shows the physical differences between these groups to be much less significant than is popularly supposed.

The Negroid race has relatively long arms and legs, woolly hair, relatively little hair on the face and body, full lips, and a flat nose. Skin color is black or dark brown. This stock has been subdivided into the African Negro proper, the Oceanic Melanesian, the Pigmy Black, and the Bushman.

There still remain certain groups of people of doubtful classification. These cannot be assigned definitely to any one of the three major human stocks. These peoples exhibit the characteristics of more than one group and would thus be classified inconsistently with regard to different sets of racial criteria. They include such groups as the Australian and Indo-Australian (Veddoid), the Polynesian, and the Ainu, a people of low cultural status inhabiting an island off the coast of Japan. The Ainu have both Caucasian and Mongoloid traits, but are characterized by a thick hair-covering on the entire body. The impossibility of classifying these groups is not a serious deficiency of the present schema, however, since they comprise only a very small segment of the human species. It has been estimated that about 99% of all mankind can be assigned to one or another of the three major races.

This classification is a decided improvement over the traditional "five races," but some of its subdivisions probably still represent oversimplifications. This is particularly true of the tripartite division of European peoples into Nordic, Alpine, and Mediterranean. In an extensive analysis of available data from a variety of sources, Coon (13) concluded that the "races of Europe" fall not into three but into ten principal categories, several of which can be even further subdivided. His principal categories include: Brünn, Borreby, Alpine, Ladogan (two subdivisions), Lappish, Mediterranean (three subdivisions), Nordic (four subdivisions), Dinaric, Armenoid, and Noric. Coon has called attention to the fact that the simpler classification into Nordic, Alpine, and Mediterranean, first proposed by Ripley in 1899, may have paved the way for facile "typing" of individuals and glib use of racial catchwords. It is certainly easy to slip into the stereotype of "tall, blond Nordics," "short, dark Medi-

terraneans," and Alpines who are "intermediate" both geographically and physically. An awareness of the probable complexity of the total picture may help to check such oversimplified stereotyping.

NATIONAL AND LINGUISTIC GROUPINGS

Racial affiliation should not be confused with nationality. A race is a biological group; it implies a certain community of hereditary background and is identified by physical criteria. A nation, on the other hand, is a cultural, political, and geographic grouping. It has been a common practice, especially in the popular literature on the subject, to regard all the individuals of a given nation as members of a single race. This is far from the truth and can yield only misleading results.

In France, for example, can be found Nordics, Alpines, and Mediterraneans (as well as other groups included in the finer classification discussed above), different strains predominating in different parts of the country. In modern Germany, true Nordics are relatively scarce. Certain Nordic sub-groups predominate in small areas of Germany, but other regions are populated principally by Borreby, Ladogan, Alpine, Noric, and Dinaric strains. Although Mediterraneans predominate in southern and central Italy, northern Italy is largely Alpine and Dinaric. Mediterraneans also predominate in certain sections of Ireland, England, Scotland, and Wales, other regions containing principally Nordic, Brünn, and a few other scattered strains. It is thus apparent that any racial classification must be made on an individual rather than on a national basis.

Another common source of confusion is that between racial and linguistic or philological categories. Thus such terms as "Latin," "Aryan," and "Semitic" are frequently employed in popular discussion to signify races. But the groups which now speak languages of Latin origin—including French, Italian, Spanish, Portuguese, and Roumanian, among others—present an extremely varied racial composition and are not a unit in any but the philological sense. The term "Aryan" is likewise a very broad one applied by students of linguistics to all those peoples using a derivative of the original Indo-European language. Similarly, the term "Semitic" refers to a group of languages and not to any biologically distinct group of people.

In this connection, mention may also be made of the Jews, a group characterized by religious and other cultural uniformities, but racially very heterogeneous. Believed to be originally Mediterranean, this group now contains Nordic and Alpine elements, as well as mixtures of several other European strains (13). Alpine characteristics, such as brachycephaly, are more common than Mediterranean characteristics in the group today, but the most conspicuous fact is undoubtedly the wide diversity of physical types represented.

The loose use of national, linguistic, and even religious nomenclature interchangeably with racial designations has further complicated an already difficult problem of classification. It is well to bear in mind the distinction between these various types of categories.

RACE MIXTURE

An additional difficulty in the way of racial classification is introduced by the extensive amount of race mixture which has been going on for countless generations. Such mixture is particularly common among the sub-groups of the white race. Consequently, it is difficult to find many "pure" Nordics, Alpines, or Mediterraneans even in those regions which are supposed to be characteristically populated by these groups. Similar interbreeding has occurred to a greater or lesser extent among nearly all racial groups. There exist at present only a very small number of isolated primitive groups which may be regarded as racially "pure."

When the racial mixture has occurred in violation of social dictum or group mores, as in the case of whites and Negroes in the United States, the problem of racial identification is further confused by the arbitrary classification imposed by society. Thus a "Negro" in many parts of the United States means an individual with any discoverable traces of Negro ancestry. Biologically such an individual may be much more closely affiliated with the Caucasian than with the Negroid race, but culturally he is usually a member of the Negro group and shares the social heritage of the latter.

Race mixture, or *miscegenation*, is a problem which has aroused much discussion in its own right. Its advantages and disadvantages have been argued at great length; enthusiastic exponents can be found for both sides of the question. Among those who consider miscegenation biologically injurious may be cited Davenport (15)

who argued that race mixture produces physical as well as psychological "disharmony," the mixed group being a "badly put-together people." Negroes, for example, have relatively long limbs, whites relatively short limbs. Interbreeding between these two groups might, according to Davenport, result in individuals with long legs and short arms, or vice versa. Similarly, the mixture between a race with large teeth and large jaws and one with small teeth and small jaws might produce individuals with disproportionate combinations of jaws and teeth. This was, in fact, offered by Davenport as a possible explanation for the frequency of tooth decay in the United States, since Americans represent a mixture of so many different strains!

The fallacy in this argument lies in its assumption that specific organs are inherited as unit characters. The relation between an individual's bodily or psychological traits and his gene constitution is, of course, much more complicated than that. In the process of growth, moreover, all parts of the organism interact and influence each other's development, thus producing a balanced and harmonious relationship of parts.[4] Observations on hybrids, both in the human and in infrahuman species, reveal no significant disharmonies. The success of many animal breeding experiments certainly testifies to the beneficial results which may be obtained with race crossing. Physical measures of hybrid races have likewise shown either an increased physical vigor in the hybrid generation or a physical status which is midway between those of the parent groups. *In no case has a consistent physical inferiority of a hybrid group been reliably established.*

The effects of race mixture have also been discussed from the standpoint of the *historical achievements* of various groups (cf. 48). Two opposed theories have been proposed regarding the influence of race mixture upon the rise and fall of civilizations. On the one hand are cited ancient Egypt, classical Greece, and the Roman Empire, whose decline coincided with a widespread intermixture with culturally underprivileged immigrant or servile groups. Similarly, the relative backwardness of certain present-day groups, such as are found in Mexico and South America, has been attributed to the fact that they are of hybrid stock.

An equally strong case can be presented, however, in support of the opposite theory. Racial purity is often associated with a very low level of cultural development. Thus among the most racially

[4] For a fuller discussion of these criticisms, cf. 10, 11.

pure human groups may be mentioned the hill folk of India, the Andaman Islanders, and certain Eskimo groups. In our own country, the closest approximation to purity of racial stock is probably to be found in certain isolated mountain communities, which are notoriously backward in social and intellectual development. Conversely, the achievements of western civilization can be shown to be the cultural expression of hybrid stocks. All the great European nations present a complexity of racial composition. The history of the United States furnishes a particularly striking example of the achievements of a highly mixed group. It can also be shown that many great men were the product of much interbreeding of diverse stocks.

The apparent inconsistencies in such data arise from the attempt to establish a *causal relationship* between race mixture and cultural level. There is, in fact, no reason to expect a direct relation between the two. Both may in turn be dependent upon a third factor, the degree of social contact or social isolation of a group. Cultural development is usually promoted by contacts between groups, with the resulting interchange of diverse material and intellectual products. At the same time, such contacts are conducive to race mixture. Hence a heightened cultural development is often found in association with race mixture.

In certain situations, social factors may cause the reverse relationship to hold between degree of racial purity and cultural development. Thus in a period of cultural degeneration, miscegenation with a despised group may be tolerated as social barriers are lowered. In such a case, as in ancient empires in their decadent periods, the race mixture is but another symptom of a disruption of traditional behavior and may temporarily coincide with a period of low intellectual achievement and cultural deterioration. In either case, the association between race mixture and cultural level is an indirect one, and cannot be cited as evidence for a biological basis of cultural development.

IMMIGRANT GROUPS

Many alleged "racial" comparisons have been made on immigrant groups in the United States, the subjects being classified according to country of birth. If American-born children of immigrants are employed, they are usually classified on the basis of parents' birthplace.

Such investigations cannot yield any information on the problem of *race differences.* As has already been pointed out, national groups cannot be assigned *as a whole* to one or another racial stock. But even for the study of *national differences* such data are inadequate and misleading. Immigrants cannot be assumed to be representative samplings of their nome population. They are not drawn proportionately from all educational, economic, and social levels, but usually represent a select group.

A more serious difficulty is that such selective factors may operate differently in each country. As a result, immigrant groups from different nations are *neither fair samplings of their home populations nor comparable among themselves.* If it could be shown, for instance, that immigrants from all nations were drawn consistently from the lower socio-economic levels, then such groups would at least be comparable with each other. But it is well known that, through purely historical reasons, the immigrants from some nations may represent a relatively inferior sampling of their population, from others a more nearly random or average sampling, and from still others a relatively superior sampling. Moreover, the nature of the sampling from a given country may change markedly from time to time.

It has been frequently suggested, for example, that the superior performance of Chinese and Japanese children in America on many of our intelligence tests may be the result of selective factors, only the more progressive families emigrating from these countries (cf., e.g., 14, 49). Many of the immigrants from southern Europe, on the other hand, are probably an inferior sampling of their own national population. In one investigation (22), groups of Danish and Italian girls in the United States and in Europe were examined with the International Group Mental Test. Although the Danish samplings in this country excelled the Italian, no significant difference was found between the groups tested in Copenhagen and in Rome.

It is apparent that the testing of immigrants can throw little or no light upon the relative status of the national groups from which they are drawn. It might be argued, however, that the determination of the abilities and personality traits of the immigrants themselves is of direct practical value for restriction of admittance, assimilation, and similar purposes. Such an argument fails to take into account two important aspects of the problem. In the first place, the behavior of immigrants may simply *reflect their former environmental background.* We

cannot assume that the emotional and intellectual traits of such persons are innately determined just because they persist under the new environment. The influence of early conditioning is too strong to be readily wiped out. Similar traits would also be noticeable to a slighter extent in the offspring of immigrant parents, as long as family traditions and the practices of the home country endure.

A further point to note in the study of immigrant groups is that the immigration itself, with its resulting necessity of *adjusting to a new culture,* is an important environmental influence (cf., e.g., 52, 58). This factor cannot be ignored in analyzing the intellectual and emotional make-up of the immigrant. The confusion of standards and the shifting reference points contingent upon such an adjustment cannot fail to have an effect upon the subject's behavioral development. The point has frequently been made that the maladjustment is greatest, not in the case of the immigrating generation who retain their customs to a large extent, nor in the case of the third and succeeding generations where adaptation and assimilation is virtually complete, but in the case of the offspring of the immigrants—or second generation—who are caught in the maelstrom occasioned by two different frames of reference.

For example, a survey with the Woodworth-Mathews Test of Emotional Instability revealed a much higher average number of neurotic symptoms among the children of immigrants than among those of native parentage (45). The children tested ranged in age from 9 to 19 and in school grade from the fourth to the twelfth. Both sexes were included in the study. The median number of neurotic symptoms reported by each of the three major groups selected for comparison was as follows:

"Mixed" group: largely of north European ancestry; resident in America for several generations (N = 87)	16
Jewish group (N = 199)	20
Italian group (N = 188)	36

Data such as these do not constitute an adequate basis for the conclusion that Jewish or Italian groups in this country are by nature emotionally unbalanced. In a similar situation, the "normal" individual upon whom the test was standardized might have reacted similarly. The fact that immigrant groups often live under poorer *socio-economic conditions* than the native population may likewise affect

their intellectual, emotional, and social adjustment. Such effects will be considered in the appropriate section of the following chapter.

DIFFERENTIAL SOCIAL SELECTION

A type of selective sampling which complicates certain comparisons among racial or national groups results from differential social selection. We have already seen examples of such differential selection as it operates with respect to men and women (Ch. 18). In a similar manner, it may operate with respect to various minority groups living within the same country. In evaluating any results on special groups, such as college students, army inductees, or institutional populations, we need to be on the alert for possible spurious effects resulting from such selective factors.

For example, comparisons of the test performance of Negro and white soldiers in World War II are complicated by the fact that Selective Service screening standards were apparently different for the two groups (32). Similarly, in World War I, the policy with respect to the administration of the Alpha and Beta examinations varied somewhat from camp to camp because of practical exigencies (cf. 24). Thus in some localities men who scored below 30 on Alpha might be re-examined with Beta. In others, because of more demand for Beta-testing, all those who obtained an Alpha score barely higher than zero were classified on the basis of Alpha alone. This would obviously affect the comparisons of Alpha or Beta averages among groups tested in different localities. Since the proportion of foreign-born or of Negroes also differed from camp to camp, comparisons between these groups and native-born white draftees would be correspondingly affected. As a matter of fact, such a practice would tend to *exaggerate* the differences among the groups being compared. Those groups in which the need for Beta was more prevalent (because of language handicap, illiteracy, and the like) would be the very ones in which the use of Beta had to be restricted more stringently.[5] Hence in such groups only persons receiving Alpha scores close to zero would have the benefit of a retest with Beta.

Another illustration of differential selective factors is provided by comparisons of Negro and white college students. In one typical investigation (21), Negro college girls were found to be significantly more

[5] This was not, of course, true in all camps, but only in those in which testing facilities were relatively inadequate

"self-sufficient" as indicated by the Bernreuter Personality Inventory, the remaining scores on this test yielding no significant differences between the two groups.[6] Does such a finding demonstrate that Negroes are more self-sufficient than whites? Obviously not, since only college girls were tested. Does it indicate that in the upper intellectual levels Negro girls are more self-sufficient than white girls? Not necessarily, because Negro girls who go to college may represent a selected sampling not only with respect to intellectual level, but also with respect to a number of personality traits. It may *require* more self-sufficiency for a Negro girl to continue her education than it does for a white girl, because of the relatively greater economic and social obstacles which the former must overcome. Any personality difference between the two groups may thus do no more than reflect these differences in the operation of selective factors. Of course, we must also consider the possibility that going to college may itself be a factor in increasing the self-sufficiency of a Negro girl. The realization that one has successfully surmounted obstacles is probably an important condition in the development of feelings of self-sufficiency.

Statistics on crime and insanity are especially subject to differential selective factors. Statements have often been made regarding the "predisposition" of various racial groups to crime. The large percentage of crime in the United States has been attributed by some to the influx of certain classes of immigrants into our country. Statistics have been cited to show the greater frequency of crime among Negroes and among immigrants from eastern and southern Europe than among the native-born white population.

Figures often lie, and in the interpretation of crime statistics it is particularly difficult to disentangle the many uncontrolled factors which confuse the issue. Among such factors may be mentioned the inequality in arrests and convictions among various groups; Negroes and "foreigners," for example, are more readily arrested "on suspicion" and on less grounds than is generally required for native-born whites. The fact that most foreigners are adults would also give them a disproportionate percentage of crime if they are compared with the figures for native-born persons of *all* ages. Similarly, foreigners are more often city-dwellers and live under poorer social and economic conditions than native-born Americans—all of which is conducive to

[6] Cf. p. 672 in Chapter 19 for a list of the other aspects of personality covered by this test.

crime. The foreigner, furthermore, may have brought with him traditions and folkways which happen to conflict with the accepted behavior in our country. Mexicans in the United States, for example, show a relatively large number of arrests for carrying concealed weapons (60). In this they may simply be continuing habits which they acquired in their own country in a perfectly legitimate way. Despite the many factors which load the dice against the foreign-born in crime statistics, careful analyses of the data on native and foreign-born persons *over 18 years of age* have failed to reveal a higher rate of arrests, convictions, or commitments among the latter (cf. 60).

The American-born children of immigrant parents present a somewhat different problem. On the whole, the crime rate among such "second-generation" Americans is higher than among offspring of native parents (28, 52). The conflict between the old and new culture is undoubtedly an important factor in the emotional and social maladjustment of such individuals (27, 51, 52). At the same time, it should be noted that foreign parentage need not in itself be associated with a higher crime rate. Surveys have shown that in many states the sons of immigrants have a lower commitment rate than the sons of native parents (9, 55). Those states having a higher crime rate among persons of foreign parentage are generally the more highly industrialized and urbanized states. They also contain a larger proportion of immigrants from those European cultures which are most unlike our own. Thus the socio-economic conditions under which the immigrant groups live, as well as the degree of conflict between the old and new cultures, seem to be major factors in determining the crime rate.

Most of the conditions which render the evaluation of crime statistics difficult also affect the data on insanity. In addition may be mentioned the factor of hospitalization. Institutional subjects may not be a representative sampling of the actual cases of mental disorder in different groups, since the available facilities for hospital care are not equal in different parts of the country. On the other hand, because of economic conditions, certain groups are better able to care for the mentally disordered persons at home, thus eliminating the necessity for hospitalization. It is interesting to note that, although the uncorrected hospital statistics show about twice as many cases of insanity among the foreign-born as among the native-born, the difference virtually disappears when various corrections are made for sampling inequalities (42).

CHAPTER

21

Racial Comparisons: Problems of Measurement[1]

IN THE PRECEDING CHAPTER it was shown that the classification of individuals into distinct races, as well as the choice of groups suitable for racial comparisons, presents many difficulties. Even when a satisfactory selection of subjects has been made, however, additional problems remain to be solved. It is not sufficient to determine *whom* to measure. The questions *what* to measure and *how* to measure it are equally important. Thus it is necessary to decide which are the most significant traits for comparison and what materials and techniques are applicable to the testing of culturally dissimilar groups. The interpretation of the obtained differences also raises important questions. Is it possible to establish a universal criterion of "intellect" so that we may speak of one group as being intellectually "superior" and another "inferior"? What shall we use as norms or standards for the evaluation of widely diverse peoples? The latter is a very fundamental issue in differential psychology.

Individuals who differ in racial affiliation also differ in many other respects. It is therefore very difficult to *isolate* the factor of race so as to determine its direct influence upon the subject's behavioral development. Members of different racial groups frequently speak different languages, a fact which greatly restricts the range of traits in which inter-group comparisons can validly be made. The differences in general educational opportunities and specific type of training available to each group have an undoubted influence upon psychological test performance. Such groups may likewise differ in their general social and economic level and in the facilities for intellectual advancement

[1] For the sake of brevity, the term "race" will be employed without quotation marks or other qualifications to refer to groups so designated in the particular investigation under consideration. It is not to be assumed, therefore, that such groups constitute races in the sense in which this term was defined in the preceding chapter. In each case, the nature of the groups will be apparent from the context.

offered in their own homes. The background of tradition and culture against which the individual develops is also fundamentally diverse from group to group. The emotional attitudes, interests, ideals, and preferences fostered by such surroundings will not be the same. To this may be added the many difficulties arising when an examiner from one racial or national group administers psychological tests to subjects in another group. This situation is not comparable to the testing of subjects within one's own group.

A considerable body of evidence is available which demonstrates the influence of the above factors upon psychological test performance. Frequently such data were gathered incidentally in studies whose major purpose was the establishment of race differences in ability. A few investigations, on the other hand, have been conducted with the explicit aim of analyzing the pitfalls in racial comparisons. In either case, the data seem clear in their implication that factors other than race are operative in alleged racial differences. It should be noted that the question of race *versus* culture in the production of group differences is one phase of the general problem of heredity and environment. Race, it will be recalled, is a biological concept based upon hereditary community. Culture, on the other hand, refers to the environmental conditions and behavior shared by the members of a single group.

THE COMPARATIVE ACHIEVEMENTS OF DIFFERENT RACES

The point is sometimes made that the vast differences in the achievements of various races testify to their dissimilar innate psychological equipment. Thus it is argued that the differences in cultural level among racial groups might be a *result* rather than a cause of psychological differences among such groups. The cultural milieu in which the individual is reared, with its special opportunities and limitations for intellectual and emotional development, might itself be a reflection of the capacities of each race. The individual of a given race might thus be handicapped by poor facilities for intellectual development just because his predecessors lacked the capacity to produce a more "favorable" environment.

As evidence of the wide inter-racial variations in achievement are cited contributions to science and invention, accomplishments in the realm of literature and the other fine arts, complexity of social and

political organization, technological advances, and many other aspects of cultural status. Comparisons have also been made in terms of the "eminent" men produced by each race. Thus Galton (24, pp. 325–337) at one time proposed a 16-point scale for estimating the "comparative worth of different races" by comparing the relative merits of men in each group who had achieved eminence. On this basis he suggested, for example, that the Negro is two grades lower than the Englishman, and the modern Englishman two grades below the Athenian of Greece's golden era.

It should be noted that any argument based upon the relationship between the cultural level and the capacity of races is reversible. On the basis solely of the association between these two factors, it is impossible to determine which is cause and which is effect. There is therefore no reason for concluding *ipso facto* that racial differences in cultural achievement indicate or result from a racial difference in capacity. There is considerable evidence, on the other hand, which suggests that the cultural differences may be responsible for the group differences in "capacity."

In the first place, achievement and cultural level are frequently found to vary not with race but with environmental factors. Thus a group which is characterized by a given achievement level may be racially very heterogeneous and may constitute a unit only in terms of a common experiential background.[2] Secondly, the relative achievements of a given group are influenced by a number of factors which cannot themselves be attributed to racial capacity without stretching the point unduly. The characteristics of the physical environment, the degree of contact with other groups, the discovery of new routes of travel and communication, and historical events within *other* groups—and thus not within the control of the group under consideration—have played an important part in the cultural development of many societies.

Thirdly, mention may be made of certain broad shifts in the relative cultural status of different racial groups from time to time. This is particularly well illustrated by some of the ancient African civilizations, such as the kingdom of Benin, whose achievements in many fields far outstripped the European cultures of the same period. A number of "lost arts" of these civilizations represent, in fact, abilities which have never been attained in any other group. In several cases,

[2] Data bearing on this point will be found in the following chapter.

the shifts in relative cultural level occurred in the absence of any known change in the nature of the stock, as might occur through race mixture. Concomitant historical developments can, however, be found which might account for the change in cultural level. Finally, the reader may consider in this connection the weight of the evidence from other sources, discussed throughout the present book, which indicates the extent to which behavioral development depends upon environmental factors.

In connection with the comparative achievements of different races, mention should also be made of the theory that "primitive" man excels in sensory capacities, in contrast to "civilized" man's "superior intellectual equipment." This theory has been especially proposed as an explanation of the remarkable feats of primitive persons in such tasks as the recognition of birds or animals concealed among foliage, the interpretation of footprints, the use of olfactory cues in finding one's way or in identifying animals, and the like. Considerable evidence has been accumulated, however, to show that such achievements are not attributable to superior sensory equipment. Rather do they result from the individual's having *learned* to respond to very slight cues and to discriminate small differences. The situation is roughly similar to that underlying the blind person's skill in responding to auditory and tactual cues. The needs of life in a primitive environment are such as to encourage the learning of proper responses to slight sensory cues which may spell danger, food, or other urgent matters. That such achievements result from learning rather than from race differences in acuity is suggested by the fact that persons from "civilized" countries have proved able to learn similar responses when put in situations which demanded them.

Objective tests of sensory acuity have likewise lent no support to the view that primitive man's achievements result from sensory superiority. As early as 1904, at the St. Louis World's Fair, Woodworth (63) and Bruner (11) applied what few tests were then available to such groups as American Indians, Negritos from the Philippine Islands, Malayan Filipinos, Ainus from Japan, Africans, Eskimos, Patagonians, and others. White visitors to the exposition were similarly tested. On such controlled tests of sensory acuity, the primitive groups did no better than the white norms. Subsequent investigations on many different groups have corroborated these findings.

A similar explanation in terms of learning rather than sensory dif-

ferences seems to hold for alleged racial differences in musical achievements. The aesthetic intricacies of Indian dances have led many observers to ascribe a superior musical sensitivity to that race. Even more familiar is the traditional musical talent of the American Negro, whose achievements in this respect have become an important element of American music. That cultural rather than racial factors account for these accomplishments is suggested by extensive surveys with the Seashore Measures of Musical Talents. In the discrimination of pitch, intensity, time, and rhythm, as well as in the other simple tests in this well-known series, no significant superiority in favor of Indians or Negroes has been found (5, 26).

It is thus apparent that the cultural achievements of different groups cannot in themselves serve as an index of the relative abilities of human races. In the effort to obtain a more direct measure of abilities, psychologists have administered a wide variety of tests to individuals of different races and cultures. A mass of data, ranging from indices of simple sensori-motor abilities to measures of complex intellectual and emotional characteristics, have thus been accumulated on various racial groups. The *interpretation* of these findings, especially in the case of the more complex functions, is beset with many difficulties or pitfalls. In the remaining sections of the present chapter, we shall consider some of these difficulties.

LANGUAGE HANDICAP

It is obvious that in the comparison of groups speaking different languages, verbal tests cannot be employed. Non-language and performance tests have been devised for this purpose. It is not to be concluded, however, that the same traits are being measured by all these tests. As was shown in Chapter 14, many tests included under the heading of "intelligence scales" call into play widely different abilities. Thus when unfamiliarity with the language makes the application of verbal tests impossible in a given group, the range of processes which can be measured in that group is thereby narrowed. There is no substitute for verbal tests. It is a psychological impossibility to eliminate the verbal content of a test without altering the intellectual processes involved.

The actual effect of language handicap upon test performance is likely to be most serious, however, when such a handicap is present

in a mild degree. If the individual has a moderate understanding of English, it is usually deemed unnecessary to give him a non-verbal test. But such an individual may lack the facility in the use of English or the range of vocabulary required to compete fairly on a verbal test. This situation is often encountered in immigrants who have lived in America for many years, or in the children of immigrants. The latter are frequently bilingual, speaking their own language at home and English at school.

The fact that children with such relatively mild language handicaps generally obtain lower intelligence test scores has been frequently demonstrated. When children in the same schoolroom are tested with a common verbal intelligence test, those from foreign-speaking homes generally make a poorer showing as a group. In an investigation conducted in New Jersey, for example, children of Italian parentage were given the Otis Group Test (46). When the children were divided into four language groups—those who spoke only Italian at home, those who spoke Italian and some English, those who spoke English and some Italian, and those who spoke exclusively English—a consistent rise in average score was found with increase in amount of English spoken at home.

In an analysis of data secured independently by different investigators, Goodenough (30) found a correlation of −.75 between the average IQ of children in various immigrant groups and the tendency of such groups to retain their own language for use in the home. An index of the latter was obtained by finding the ratio of the number of parents in each national group who had been in this country for a period of 20 years or over and had not adopted English, to the total number of parents in that group who had adopted it. The high negative correlation between these two factors indicates a strong tendency for children in those immigrant groups in which English is not readily adopted to obtain lower scores on our intelligence tests. Goodenough points out that there are two possible explanations of such a finding. On the one hand, the lower intelligence test scores of some groups may result directly from their greater language handicap. On the other hand, those national groups in which English is not commonly adopted may be less intelligent and less progressive from the outset. Their failure to learn English would thus be the result of lower intelligence and poorer adaptability.

Neither interpretation can be selected solely on the basis of the cor-

relation between the two factors. Other data, however, suggest that the former is the more probable one. It is quite likely, for example, that because of the *greater similarity of some languages to English* it is easier for individuals from certain countries to learn English, quite apart from their intellectual level. Another factor of possible relevance is to be found in the reasons for which immigrants from various countries come to America. Those from some countries may come largely with the intention of settling permanently; those from other countries may traditionally retain a vague hope of returning to their home country after "making their fortune." Such *impermanence* is likely to be reflected in their halfhearted attempts to master the English language or to see that their children master it. A further question is whether the immigrants come into a *community of their own compatriots*, as found in the foreign neighborhoods of some of our larger cities, or whether they are scattered in predominantly American communities. In the case of certain national groups, represented by relatively small numbers of immigrants and not concentrated in any one area, the individual has little choice but to learn English.

The most crucial argument regarding the role of language handicap, however, is provided by the finding that *the inferiority of the immigrant groups is greatly diminished and may disappear entirely when non-language tests are employed.* This has been repeatedly and consistently demonstrated with many groups. Pintner (53) tested 165 school children of Italian, German, and Polish parents and 121 children of American-born parents, mostly of Irish descent. On the National Intelligence Test, a predominantly verbal test, only 37% of the foreign-parentage children reached or exceeded the median score of the native-parentage children. On the Pintner Non-Language Scale, however, 50% of the foreign group reached or exceeded the native median, i.e., the two groups had identical medians. A similar result was obtained in an investigation at the preschool level (16). Two groups of nursery school children, one bilingual and the other monoglot, were matched in age, sex, and paternal occupation. The second language was Italian in all cases. The two groups, each consisting of 106 children, were given the 1937 Stanford-Binet and a performance test, the Atkins Object-Fitting Test. The bilinguals were significantly inferior to the monoglots on the Stanford-Binet, but significantly superior to the monoglots on the object-fitting test.

In a carefully controlled and extensive study, Arsenian (2) admin-

istered the Pintner Non-Language Scale to 1152 American-born children of Italian parents and 1196 American-born children of Jewish parents. The children ranged in age from 9 to 14. Degree of bilingualism was ascertained by means of a written questionnaire. The results showed no significant correlation between extent of bilingualism and intelligence test score in either group. The correlations were —.079 and —.193 in the Italian and Jewish groups, respectively.

It is interesting to note that, when the language handicap is fairly pronounced, even the use of English in giving directions in a non-verbal test may affect the scores. In one study, 236 Spanish-speaking children in the first three grades of Arizona public schools were given the Otis Primary Group Intelligence Test (48). This is a non-verbal test with oral instructions. To check the effect of bilingualism, one half of the group was given Form A with Spanish instructions, followed about ten days later by Form B with English instructions. The procedure was reversed in the other half of the group. The mean IQ was found to be 96.15 on the Spanish form and 86.87 on the English form. No child received a higher IQ on the English than on the Spanish form, although 9 received the same IQ on both. The largest individual difference in favor of the Spanish form was 44 points. It should be noted that this was a non-verbal test, Spanish being used only to give directions. Had the children been given a predominantly verbal test in Spanish, they might have done just as poorly as they would on a verbal test in English. When a child speaks one language at home and another at school, his mastery of *both* languages may suffer as a result.

This was demonstrated in an investigation conducted on Welsh children (4). Two verbal intelligence tests, a word-knowledge test, and the Pintner Non Language Scale were given to 10- and 11-year-old children in a Welsh-speaking and in an English-speaking district of Wales. School instruction was conducted in English in both areas. As in other studies, the bilingual children were found to be superior on the non-language scale but inferior on the verbal tests. When Welsh forms of the verbal intelligence tests and the word-knowledge tests were administered, the inferiority of the bilinguals was even greater than it had been on the English forms. In other words, these children did even more poorly in their "home language" than they did in their "school language." Such a result is not, to be sure, surprising. If we consider the acquisition of vocabulary as an example, it is apparent

that when the child speaks one language at home and another at school, he will learn a somewhat different set of words in the two situations and his vocabulary in *each language* will thereby be curtailed.

A further point to bear in mind is that bilingualism *per se* does not necessarily result in language handicap. Under certain circumstances, bilingualism may produce no handicap in one or both languages, and in such cases we would not expect it to depress intelligence test performance. For example, Pintner and Arsenian (54) found no relationship between degree of bilingualism and scores on a verbal intelligence test in a group of 469 native-born Jewish school children in New York City. The correlation was -.059 and fell well within the chance value. It has been a general finding, in fact, that Jewish school children, as well as college students, do especially well on verbal tests. Thus the bilingualism of the Jewish child is not such as to interfere with his mastery of English. One reason for this may be found in the attitude of the Jewish group toward the two languages, as contrasted with the attitude of other foreign-language groups. The Jewish child in America will eventually have to make his way in an English-speaking society, and English is therefore of prime importance to him. On the other hand, those national groups which are in large part oriented toward the possibility of returning to their country of origin may regard English more as a temporary expedient. Another important factor is undoubtedly the strong educational tradition in the Jewish culture and the parental insistence that the child do well in school, especially in the relatively abstract and "bookish" subjects (8; 39, p. 174).

Another illustration of the fact that the mere acquisition of a second language need not prove to be a handicap is provided by an investigation conducted in Ireland (59). All the children tested were predominantly English-speaking at home, but some attended Irish-speaking schools and some English-speaking schools. In both types of schools, however, the second language was taught as a school subject, i.e., English was taught as a separate course in the Irish-speaking schools, and Irish was so taught in the English-speaking schools. On performance tests, both groups did equally well. On verbal tests, however, the children in the Irish-speaking schools excelled those in the English-speaking schools *when tested in English*. When the tests were administered in Irish in the Irish-speaking schools and in English in the

English-speaking schools, the children in the English-speaking schools obtained higher scores. It should be remembered that all children were bilingual to a certain extent in school, because of the course instruction in the second language. What these results actually show is that children will do better on an intelligence test administered in the language which they speak at home *and* study as a school subject than they will in a language which is used in school instruction but not encountered elsewhere. The data further suggest that, when children are bilingual, they master more thoroughly a language which they speak at home and learn as a school subject than one which they merely speak at home and in school.

In summary, bilingualism *per se* need not handicap a child. Bilingualism as it occurs in a large proportion of the immigrant population, however, is such as to reduce the child's mastery of *either* language, because one language is restricted largely to one set of situations in the child's life, and the other language restricted to another set. What the child needs is to learn to express himself in at least one language in all types of situations. It is not the interference of the two languages, so much as the restriction in the learning of one or both to limited areas, that produces a handicap.

TABLE 44 *Median IQ's of Indian School Children*

(Adapted from Jamieson and Sandiford 37, pp. 540-542)

Test	*Number of Cases*	*Median IQ*
National Intelligence Test	275	79.8
Pintner Non-Language Test	280	96.9
Pintner-Paterson Performance Tests	115	96.4
Pintner-Cunningham Primary Test	59	77.9

Studies on language handicap have not been limited to European national groups. In investigations on American Indians, the influence of language deficiency upon intelligence test performance has been vividly demonstrated. Jamieson and Sandiford (37) administered a series of standard tests to 717 pupils attending Indian schools in Ontario, Canada. All the children could speak English, but their ability to do so was below that of the average American child. The median IQ's obtained by the Indian children on each test are shown in Table 44.

A comparison of the median IQ on the verbal and non-verbal tests reveals the influence of language handicap. On the National Intelligence Test, a predominantly verbal test, the Indian children are clearly below the American norms. On the Pintner Non-Language and Pintner-Paterson Performance Tests, on the other hand, their performance is practically up to the norms.[3] The low median IQ on the Pintner-Cunningham Test, administered to the younger children, again suggests the role of language handicap. Although non-verbal in content, this test requires extensive and detailed instructions given in oral English.

A more conclusive demonstration of the importance of language handicap was provided in the same study by a comparison of a group of monoglots, who spoke only English, with bilinguals, who spoke an Indian language at home all or part of the time. The median IQ's of these two groups on each test are shown in Table 45. It will be noted

TABLE 45 *Median IQ's of Monoglot and Bilingual Indian School Children*

(Adapted from Jamieson and Sandiford, 37, pp. 540 542)

Test	*Number of Cases*		*Median IQ*	
	Monoglot	Bilingual	Monoglot	Bilingual
National Intelligence Test	153	115	82.4	76.6
Pintner Non-Language Test	152	121	100.0	93.6
Pintner-Paterson Performance Tests	80	30	95.8	100.0
Pintner-Cunningham Test	33	23	80.5	68.1

that *on the performance scale the bilingual children obtain a higher median IQ than the monoglots, whereas the reverse is true on the other three tests.* This suggests that the poorer showing of the bilinguals is not due to their inferior mental status but to the verbal nature of the test. In the case of the Pintner Non-Language Test, it is possible that the use of paper-and-pencil materials, as well as the dependence of some of the sub-tests upon information characteristic of our culture, gave a disadvantage to the children from the less highly assimilated homes. Those children who were relatively unfamiliar with such

[3] The fact that the medians on these two tests still fell below 100 may be explicable in terms of speed of work, motivation, differences in general information, and other factors to be discussed in subsequent sections of the present chapter.

materials would also tend more often to come from Indian-speaking homes.

Subsequent investigations on a number of different Indian groups have corroborated the findings of Jamieson and Sandiford. On verbal tests, Indian children average consistently below the white norms, but on non-language and especially on performance tests they are approximately equal to white children (3, 28, 32). In an extensive survey by Havighurst and Hilkevitch, 670 American Indian children between the ages of 6 and 15 were tested with a short form of the Arthur Performance Scale (32). The subjects included Hopi, Navajo, Zuni, Zia, Papago, and Sioux. Although there were large differences among various Indian communities, the total average coincided with the white norms. A sub-group of 30 children who had also been tested with the Kuhlmann-Anderson received a mean IQ of 82.5 on the latter test and 102.8 on the performance scale. The differences in score among communities were shown to be at least in part associated with the degree of contact with white culture. For example, two groups belonging to the same tribe showed a difference in test score which corresponded to the difference in the extent of their contact with white culture.

Similar results have been obtained with Oriental groups in America. Darsie (17) tested 570 American-born Japanese children between the ages of 10 and 15. Only those children who reported that English was the language most familiar to them were included in this group. The linguistic difficulties were therefore not very pronounced, but were just such as might be commonly found among the children of immigrants. On the Army Beta, a non-language test, there was no consistent difference in score between Japanese and American children. The direction of the difference varied from one sub-test to another; the total scores showed no significant difference at ages 10 and 11, beyond which ages the Japanese children excelled.

The Stanford-Binet, however, yielded clear-cut differences. The median IQ of the Japanese group was 89.5, as contrasted with 99.5 for white children of the same districts. That this difference was attributable to the verbal nature of the test was definitely demonstrated by a special analysis conducted by Darsie. Each individual test on the Stanford-Binet scale was ranked for degree of Japanese inferiority. The tests were then rated independently by seven psychologists on the basis of the degree to which success on each depends upon verbal

ability. A final ranking of the tests was obtained by taking the average of the ratings by the seven judges. When these two sets of ranks—the one for Japanese inferiority and the other for "verbality"—were compared, they were found to correlate +.87. Further corroboration of this relationship was furnished by a comparison of the performance of Japanese and whites on the separate tests. Thus the superiority of the American children was found to consist chiefly in their greater success on the linguistic tests. The Japanese surpassed the whites, on the other hand, in certain non-verbal tests of the Stanford-Binet scale involving sustained attention and visual perception.[4]

DIFFICULTIES OF TEST ADMINISTRATION

In addition to language handicap, other special difficulties are encountered in the attempt to administer tests to widely differing groups. The use of *pantomime* and gesture in non-language tests is often confusing to the subject since it is not his normal mode of communication. This is illustrated in certain observations regarding the administration of the Army Beta to the Negro draft during World War I. Several examiners called attention to the fact that it was difficult to keep up their subjects' interest in the test. In the report from one camp, it was stated that "it took all the energy and enthusiasm the examiner could muster to maintain the necessary attention, as there was a decided disposition for the Negroes to lapse into inattention and almost into sleep" (64, p. 705). One of the reasons offered in explanation of this reaction was the artificiality of the situation produced by the elimination of language. It is also difficult to standardize directions given in pantomime and to insure that they shall always be repeated in identical fashion.

The use of *pictures* as test materials is also somewhat questionable, especially in cultures which provide no experience with pictorial representation in everyday life. A two-dimensional reproduction of an object is not a perfect replica of the original, it simply presents certain cues which, through the influence of past experience, lead to the perception of the object. If the cues are highly reduced, as in a simplified or schematic drawing, or if the necessary past experience is absent, the correct perception may not follow. It might be added that

[4] The Japanese children were significantly superior in the Induction, Paper Cutting, Code Learning, and Enclosed Boxes tests.

pictures of objects which are themselves unfamiliar in the cultural group to be examined are obviously unsuitable as test materials. They have, nevertheless, been included in certain non-language scales which have been employed in racial comparisons.

The so-called culture-free tests, such as the International Group Mental Test (19), R. B. Cattell's Culture-Free Intelligence Test (14, 15), and the Leiter International Performance Scale (41), make a deliberate attempt to include only content which is universally familiar in all cultures. In actual practice, however, they still fall short of such a goal (cf. 43). In so far as they employ pictorial representation, these tests may also favor certain groups unduly. Mention may likewise be made of tests designed to measure individual differences *within* cultures quite unlike our own, such as the Fiji Test of General Ability prepared by Mann (42) Such tests do not, of course, lend themselves to inter-cultural comparisons.

A further problem arises in connection with *rapport and motivation*. Accepted testing practice demands that the examiner establish rapport with his subjects. By this is meant, in general, that the subjects should be put at their ease, their interest and cooperation should be secured, and they should be made calm and comfortable before the test is begun. In other words, it is assumed that each subject will be in a condition to do his best. In an individual test a definite effort is usually made to establish rapport with the subject. With group tests, however this is much more difficult. The examiner in such a case must limit himself to a few reassuring introductory remarks and to the elimination of any obvious handicaps under which individual members of the group may be laboring.

When an examiner from one cultural or racial group administers a test to subjects in a different group, rapport is even poorer, the situation being much more strained and unnatural for the subjects than when they are tested by a member of their own group. This is particularly noticeable in the testing of American Indians and Negroes by a white examiner. An interesting illustration of the effect of the examiner-subject relationship is to be found in a study by Canady (12), in which 48 Negro and 25 white school children were given the Stanford-Binet by both Negro and white examiners. Some of the subjects in each racial group were tested by the white examiner first, and some by the Negro examiner first. In both white and Negro

groups, the mean IQ was about six points higher when the subjects were tested by an examiner of their own race.

Quite apart from any racial disparity, the presence of a stranger will in itself occasion more emotional disturbance among the members of certain cultures than it would among American city school children, who are accustomed to sudden visits from a succession of supervisors, research workers, psychologists, and others. Furthermore, the suspicion and hostility manifested by many "primitive" peoples toward strangers will necessarily affect the individual's attitude and responsiveness toward a foreign examiner.

DIFFERENCES IN SCHOOLING

It is well known that the educational facilities available to the individual vary widely from one racial or national group to another. This is apparent even if we consider only the *total duration* of schooling. In certain rural sections of the United States, for example, the school year is drastically shortened, sometimes to as little as six months. The *irregularity of school attendance* prevalent in certain groups, such as the American Indian, reduces still further the effective length of time devoted to instruction. Finally, the *quality* of the available training and the conditions under which it is obtained cannot be ignored. In general, it is just in those groups which receive the least schooling that the quality of instruction is poorest. The type of education offered in rural Negro schools of the South, for example, is far inferior to that in the average white public school. To equate years of schooling does not eliminate educational differences between such groups.

It is now generally recognized that intelligence tests are not independent of educational background. It will be recalled that such tests are often validated against school progress as a criterion (cf. Ch. 2), and that they correlate nearly as highly with tests of school achievement as they do with other intelligence tests (cf. Ch. 14). This correlation is even higher within groups whose educational background is relatively heterogeneous. For example, in a survey of about 2000 urban Negro school children in Texas and Oklahoma, a correlation of .81 was found between an intelligence test and a test of school achievement (27).

It will be recalled that a high correlation between *amount* of school-

ing and intelligence test score was found among the soldiers tested in both world wars (cf. Ch. 8). A further analysis of Army Alpha scores and academic level, for both Negroes and whites tested during World War I, is given in Table 46. Within any one group, there is a consistent rise in median Alpha score with increase in amount of education. That differences still exist when comparisons are made vertically, within a single educational class, is attributable to a number of factors. Chief among these are differences in the *quality* of education, a factor which is ignored in the system of classification here employed. Differences in the socio-economic level of the home as well as in other more general conditions may also be mentioned in this connection. But it is apparent that the differences in score are much larger when we read *across* the table than when we read *down* any one column. In other words, differences associated with educational level are of a much higher order than differences associated with race or foreign birth.

TABLE 46 *Median Army Alpha Score of Men with Different Amounts of Schooling*

(From Yerkes, 64, Part III, Ch. 10)

Group	*Elementary School*		*High School*	*College*
	0–4 Years	5–8 Years		
White native-born draft	22.0	51.1	92.1	117.8
White foreign-born draft	21.4	47.2	72.4	91.9
Northern Negroes	17.0	37.2	71.2	90.5
Southern Negroes	7.2	16.3	45.7	63.8

To be sure, correlation never proves causation. The fact that intelligence test scores rise with educational level does not in itself tell us which is cause and which effect. It can be argued that the more intelligent individual will be more successful in his school work and will pursue his education further than the less intelligent. Intellectual differences may thus be the cause rather than the effect of educational differences. In such a case, persons in the higher educational groupings would represent a more highly selected sampling from the outset. This explanation, although partially applicable to individuals within a group, appears far-fetched when applied to different racial and national groups. *When opportunities for continued education or for satisfactory instruction at any level are so unlike from one group to*

another, failure to obtain such education cannot be attributed to inferior intelligence.

The effect of educational handicap upon intelligence test performance is especially apparent in the American Negro. Since his native language is English, the Negro is frequently tested with the common verbal type of intelligence test. Because of his pronounced educational deficiency, however, the Negro has a very limited command of the language, as well as serious gaps in other fields of knowledge. Both of these factors would seriously alter the interpretation of scores on common intelligence tests.

Of interest in this connection is the finding that 87% of the Negro and 84% of the white soldiers assigned to the army's Special Training Units during World War II completed the training successfully (61). This course consisted of an intensive educational program designed to bring illiterates up to a fourth grade level in reading and arithmetic and to give them a minimum degree of proficiency as soldiers. Some completed the course in as little as three weeks; others required as much as thirteen or sixteen weeks, but the average duration was eight weeks. These data do not, of course, have any bearing upon Negro-white *differences* in learning ability, since individuals were *selected* for admission to this program on the basis of promise (22, 62). It was principally those who had been deprived of adequate educational opportunities who were assigned to the special training. The similarity in per cent of successes among Negroes and whites simply means that the prognostic indices employed to select individuals were about equally effective in both groups. What the results actually show is that, through an intensive educational program, large numbers of individuals of both races were able to make remarkably rapid progress in the type of functions measured by intelligence tests.

SOCIO-ECONOMIC LEVEL

It is apparent that the economic, social, and cultural level of the homes of such groups as immigrants, Negroes, or American Indians is on the whole far below the general American average. One of the first investigations designed especially to determine the relative contribution of socio-economic differences and of racial or national differences to IQ was conducted by Arlitt (1). The Stanford-Binet was administered to 191 American children of native-born white parents, 87

children of Italian immigrants, and 71 Negro children. All the subjects were taken from a single school district and all spoke English with no apparent difficulty. Each child was classified on a 5-point scale on the basis of father's occupation, which was taken as an approximate index of the socio-economic level of the home.

The median IQ of each group is shown in Table 47. When the three "racial" groups are compared as a whole, the children in the immigrant and Negro groups fall 21 to 23 points below the group of native-born white parentage. The differences in occupational level among these three groups were, however, very large. Over 90% of the Italians and Negroes fell into the semi-skilled and unskilled categories. When only the children in these two occupational levels are included, the median IQ of the native white group drops to 92.0. Thus the intellectual differences among the three racial groups are reduced to a very small quantity when comparisons are restricted to children of roughly the same socio-economic level.

TABLE 47 *Socio-Economic versus Natio-Racial Factors in Relation to Children's IQ's*

(From Arlitt, 1, pp 181–182)

Group	*Median IQ*	
	Total Groups	Groups of Roughly Comparable Occupational Level
Native white parentage	106.5	92.0
Italian parentage	85.0	85.0
Negro	83.4	83.4

It might again be objected that we cannot determine which is cause and which is effect in the relationship between intellectual and occupational level. Since, however, the opportunities for employment in higher positions are far from equal for native Americans and immigrants, and this difference is still greater when Negroes are considered, it seems unwarranted to attribute the lower occupational status of the latter groups to inferior intelligence.

Several investigations on the relationship between socio-economic factors and IQ among Negroes in America have contributed toward a clarification of the nature of this relationship. First, it should be noted that the differences in intelligence test scores among occupational

classes tend to be *smaller* for Negro than for white children. Thus in a survey of third grade Negro school children in Washington, D. C., Robinson and Meenes (57) found a 13- to 14-point difference in mean IQ between the children of laborers and the children of professional men. Among white children, this difference is generally about 20 points (49). Moreover, the mean IQ's do not follow the occupational hierarchy so closely among Negro as among white children, but tend rather to fall into a dichotomy, with clerical, business, and professional occupations in the upper category, and skilled and unskilled labor in the lower (13, 52, 57). It is likely that the socio-economic level of Negro homes is less closely related to occupational class than is true of white homes. The range or heterogeneity of white homes is undoubtedly much greater than that of Negro homes. The difference between the remuneration of a Negro in business or professional work and one in the skilled trades is probably much smaller than that between whites in the corresponding occupational categories. Restricted vocational opportunities would also mean that at least some Negroes with sufficient ability and education to hold a higher level job might be engaged in lower level occupations. All these conditions would tend to reduce the differences in the IQ's of Negro children from different occupational classes.

Some corroborative evidence for these interpretations is provided by the previously mentioned study of Robinson and Meenes (57). The Kuhlmann-Anderson IQ's of 444 third grade Negro children attending Washington, D. C., public schools in 1938–39 were compared with those of 491 Negro children in the third grade of the same schools in 1945–46. In the latter year, when vocational opportunities for Negroes were better, a closer correspondence between paternal occupation and child's IQ was found. Moreover, the mean IQ of the entire group was higher in 1945–46 than in 1938–39. The latter finding may also be related to the improved socio-economic status of the second group.

That occupational level may not be as diagnostic of general home conditions among Negroes as among whites is likewise indicated by the fact that more direct measures of socio-economic level show a closer relation to Negro children's IQ's than does parental occupation. Using the Sims Score Card, which is based upon a variety of home characteristics, Oldham (50) found a consistent and clear-cut rise in Negro children's IQ's from the lowest to the highest socio-economic

levels. Similarly, Robinson and Meenes (57), in the previously cited study, report high correlations between the mean IQ of children in each of the schools in the survey and such factors as the average rental or the frequency of radios in the community. All these findings suggest that parental occupation is not as valid or adequate an index of socio-economic background for Negro as for white children.

The relation between IQ and socio-economic factors tends to rise with age in the case of Negro children (60), as it does in the case of white children (cf. Ch. 23). The interpretation of such a rise is, however, ambiguous since the tests do not measure the same functions at different age levels. In a comparison of Negro and white infants in Florida in 1931, McGraw (45) found that the whites excelled on the Bühler Babytests. The white infants, however, were also taller and heavier than the Negro infants, a physical difference which could have resulted from inequalities in prenatal and postnatal care and nutrition. That different samples of the same racial group, which are living under different physical conditions, may differ in their physical development in infancy and childhood has been demonstrated for various groups, including Mexicans (29) and Europeans (47). In Chapter 11 we have already considered the possible effects of maternal nutrition upon both prenatal and postnatal development. Differences in the rate and level of physical development may in turn affect the infant's behavior, especially the simple sensori-motor functions which predominate in psychological tests at the infant level.

A study similar to that of McGraw was conducted in 1946 by Pasamanick (51) with Negro and white infants in Connecticut. In this case the Negro infants did not differ significantly from the whites in either physical or psychological development. The investigator attributes such a finding to the fact that the Negro maternal diet in this group had more nearly approached white standards.

In investigations on American Indians, the relatively low socio-economic level of the home is an important factor to be considered in the evaluation of their test performance. Opportunities for intellectual development, as well as the general level of material comfort are far below the average for American homes. Thus, in the study by Jamieson and Sandiford (37), the homes of the Indian children received an average rating of only 13 points on the Chapman Socio-Economic Scale, as compared to the white norm of 56 points. A close correspondence between the social status of various Indian groups

and their relative standing on the National Intelligence Test was found by Garth (25).

Not only intelligence, but also emotional adjustment and other personality characteristics may be related to socio-economic factors. For example, surveys of children's play activities have shown that Negro children tend to engage in group play relatively more often than white children (40). Before attributing such a finding to a greater "sociability" of the Negro race, it is well to consider that crowded housing and inadequate facilities for many other types of play may account for part or all of such a difference.

In a personality inventory survey of 1647 children between the ages of 9 and 15, Brown (7) compared the scores of several subgroups, including: urban and rural; low, middle, and high socio-economic levels; and Jewish. Slovak, and native American non-Jewish. Adjustment scores proved to be much more closely related to socio-economic level than to either "racial" group or urban-rural residence. Statistically significant differences in such scores were found between the different socio-economic groups, but the differences among the other types of groups investigated were consistently small and insignificant. Similarly, in a survey of about 60.000 selectees examined at one induction station during World War II, the rate of rejection for various mental disorders varied with population density and with socio-economic rating of the community (35, 36). The rate also differed among various national and racial groups. but those natio-racial groups with the highest rejection rate came from communities with the highest population density and lowest socio-economic level (34).

TRADITIONS AND CUSTOMS

The particular culture in which the individual is reared may influence his behavioral development in many ways. The operation of environmental forces is not limited to the extent and quality of educational opportunities available in the school, the home, and the neighborhood. The question is not only one of amount, but of kind. The experiences of people living in different cultures may vary in such a way as to lead to basically different perceptual responses, lend a different meaning to their actions, stimulate the development of totally different interests, and furnish diverse ideals and standards of behavior.

The importance of *motivation and interest* in intelligence test performance has been repeatedly emphasized. Yet it is apparent that many of the tests in current use cannot arouse the same emotional reaction in other cultures as they do in our own. Thus for an American school child the average intelligence test bears a close resemblance to his everyday school work, which is probably the most serious business of his life at the time. He is therefore easily spurred on to exert his best efforts and to try to excel his fellows. For an Indian child, on the other hand, the same test cannot have such a significance. This type of activity has no place in the traditional behavior of his family or tribe. Similarly, many investigators have noted that among Negro children interest in intelligence tests is not as keen as among white children, and that the former seem not to be as strongly motivated as the latter.

Such differences in motivation are not necessarily limited to test situations, but may exert a broader influence upon achievement in school and in other everyday life situations. Several theories have been proposed, for example, regarding the reaction of the American Negro and other minority groups to socially imposed frustrations. Dollard (20) and Maslow and Mittelman (44) have suggested that the Negro may assume an attitude of stupidity and lethargy as a defense mechanism against frustration and oppression. According to these writers, such an attitude would provide a sort of revenge and enable the individual to avoid disagreeable responsibilities. Similarly, Brown (10) has argued that the linguistic development of the Negro may be hindered by social pressures which inhibit verbalization. Inarticulateness reduces the possibility of incurring the hostility of the dominant social group, and might thus be "cultivated" as a measure of discretion.

In addition to emotional and motivational factors, *specific local manners and social usages* may influence the subject's performance on a psychological test. Several striking examples of such traditional behavior have been reported. Thus Porteus (55), in administering performance tests to Australian aborigines, found it difficult to convince his subjects that they were to solve the problems individually and without assistance. In explanation of this behavior, he writes:

> . . . the aborigine is used to concerted thinking. Not only is every problem in tribal life debated and settled by the council of elders, but it is always discussed until a unanimous decision is reached. On many occasions the subject of a test was evidently extremely puzzled by the fact

that I would render him no assistance, especially when, as happened in the centre, I was testing some men who were reputedly my tribal brothers. This was a matter which caused considerable delay as, again and again, the subject would pause for approval or assistance in the task (55, p. 308).

Similarly, Klineberg (39) reports that among the Dakota Indians it is considered bad form to answer a question in the presence of someone else who does not know the answer. This creates a particularly difficult situation in school, where the teachers find it difficult to induce the children to recite in class. In the same group, custom forbids one to answer a question unless he is absolutely sure of the answer. The effect which this would have upon intelligence tests, in which the subject is advised to "guess" when not sure and is urged to "try his best" on a difficult problem, can be readily foreseen. The child who refuses to give any answer unless he is certain of its correctness will lose many points which he might have earned through partial credits and chance successes.

Another medium through which cultural background may influence test performance is to be found in the special *associations and meanings* which have been built up by social conditioning. In one of the sub-tests of the National Intelligence Test, the child is required to underline the two words which tell what the given item always has. One of the examples in this test reads:

Crowd (closeness, danger, dust, excitement, number) [1]

Although "closeness" and "number" are given in the key as the correct answers, it was found that among Plains Indians "danger" and "dust" or even "excitement," were frequently underscored. The experience which these children had had with crowds on the prairie had taught them that these were necessary attributes of a crowd (cf. 23).

Many other instances of such culturally determined associations can be found in intelligence test performance (cf., e.g., 39). In one of the tests of the Army Alpha Form 6, occurs the question, "Why should all parents be made to send their children to school?" Of the several alternative answers given, the "correct" one is that "school prepares the child for his later life." But this is not true for the Indian child, whose schooling often unfits him for life on the reservation. Similarly, in a sentence completion test of the National Intelligence Scale is found the statement, "—— should prevail in churches and libraries."

[1] Scale A, Form 1, Test 3, Exercise 17

The word to be inserted in this case is "silence." Among Negro children, however, this problem would be complicated by the fact that their own churches are seldom silent. Noise is not only common in their houses of worship but is frequently an integral and essential part of the ritual.

A further example of the inapplicability of a psychological test to groups differing from the one upon which it was standardized is furnished by an incident which occurred in the testing of children in the Kentucky mountains.[6] The following is one of the problems in the Stanford-Binet Scale: "If you went to the store and bought 6 cents' worth of candy and gave the clerk 10 cents, what change would you receive?" One alert young boy, upon being asked this question, replied, "I never had 10 cents, and if I had I wouldn't spend it for candy, and anyway candy is what your mother makes." Still wishing to find out if the child could subtract 6 from 10, the examiner reformulated the problem as follows: "If you had taken 10 cows to pasture for your father and 6 of them strayed away, how many would you have left to drive home?" The child now replied promptly, "We don't have 10 cows, but if we did and I lost 6, I wouldn't dare to go home." The examiner tried once more with the following inquiry: "If there were 10 children in a school and 6 of them were out with the measles, how many would there be in school?" This answer came even more promptly: "None, because the rest would be afraid of catching it too."

Finally, mention should be made of the important role of *speed* in nearly all intelligence tests and of the widely varying emphasis placed upon speed in different cultures. An investigation by Klineberg (38) on Indian, Negro, and white school boys illustrates the operation of this factor. Several of the tests in the Pintner-Paterson Performance Scale were administered to the following groups:

136 Indians attending Haskell Institute in Kansas
120 Indians at the Yakima Reservation in Washington
107 whites in rural Washington, near the Indian reservation
139 Negroes in a rural district of West Virginia
25 whites in the same district of West Virginia
200 Negroes in New York City
100 whites in New York City

[6] This incident is reported in Pressey, S. L., *Psychology and the New Education.* N. Y.: Harper, 1933. Pp. 237-238.

In *accuracy* of performance, as measured by the number of errors on each test, the Indians excelled the whites, and the Negroes were either equal or slightly superior to the whites. All measures of *speed,* on the other hand, favored the whites. A comparison of groups *of the same race but living in different environments* suggested that these differences in speed were cultural rather than biological. Thus the New York City Negroes clearly excelled the West Virginia Negroes in every comparison. Similarly, the Haskell Institute Indians were consistently faster than those tested on the Yakima Reservation. A further division of the Haskell group into those who had previously lived on a reservation and those who had lived among whites in a town or city showed the latter to excel in speed.

In explanation of these results, Klineberg calls attention to the relatively insignificant part which speed plays in the life of the reservation Indian or the rural southern Negro. Most observers are impressed with the Indian's almost complete lack of concern with speed. Time means nothing in the daily activities of the Indian. He can see no reason for hurrying through a task, especially if he finds it congenial and interesting. Thus in so far as the examiner arouses the child's interest in the test, he makes the necessity of speeding appear even more absurd. At Haskell Institute, on the other hand, time is much more important than on the reservation. The students are constantly kept busy with a variety of tasks and the entire day is carefully scheduled. The white teachers, too, foster the attitude that it is desirable to finish things as quickly as possible. Similarly, the New York City Negroes have been exposed to the hustle of life in a big metropolis, whereas the rural Negroes are adapted to a much slower tempo of activity.

THE CRITERION OF "INTELLECTUAL SUPERIORITY"

In all group comparisons, there is a tendency to go beyond the actually observed *differences* in behavior and to evaluate the *relative status* of each group in terms of some presumably universal criterion. Linear comparisons are made in terms of better or worse. Thus we frequently find national or racial groups arranged in a rank-order for "intelligence." One group is said to be "superior," another "inferior" in its mentality. Such a point of view implies either that one group

is consistently poorer than another in *all* intellectual traits, or that certain behavioral processes are universally more significant, more valuable, or even more "intellectual" than others.

Specificity of Group Differences. In regard to the first of these assumptions, it can easily be shown that racial or national groups vary in the relative inferiority or superiority which they manifest in different traits. Group differences are specific, not general. Thus Japanese children have been found to excel American children significantly in tests involving sustained attention, visual perception, or spatial orientation, while falling behind on verbal or arithmetic tests. This was demonstrated in Darsie's study (17) by the relatively superior performance of the Japanese children on four of the Stanford-Binet tests, viz., Induction, Paper Cutting, Enclosed Boxes, and Code Learning, as well as on the Digit-Symbol Learning and the Number Comparisons tests of the Army Beta. A slight superiority was also shown by these children in the Cube Analysis and Geometric Construction tests of the Beta Scale.

The relative standing of American Indian children on performance and on verbal tests has already been discussed in connection with language handicap. It will be recalled that on performance tests Indian children usually average about as high as white children. Several studies with the Goodenough Draw-a-Man Test have shown that Indian children score even better on this test than on most performance scales (18, 31, 58). A number of Indian groups have obtained higher average IQ's than white groups on the Goodenough test. In a study of boys and girls from a Hopi Indian school, Dennis (18) reports an interesting sex difference on this test, which appears to be related to cultural factors within this group. The girls received a mean IQ of 99.5, the boys 116.6. Dennis attributes this sex difference to the fact that in the Hopi culture graphic art is traditionally a masculine concern, and consequently the boys develop more interest in art and have more practice in it than the girls. A similar sex difference in Goodenough score has been observed in other Indian communities which foster such a traditional sex distinction in artistic pursuits (31).

Differences in specific traits have likewise been found in comparisons among European immigrant groups in this country. Jewish children, for example, usually excel on verbal tests and fall behind in problems dealing with concrete objects and spatial relations. In a study conducted with kindergarten children in the Minneapolis public

schools, the Stanford-Binet was administered to groups of Jewish and Scandinavian children equated in age, sex ratio, and socio-economic status (9) The Jewish children were found to be superior on tests based upon general information and verbal comprehension, while the Scandinavian children excelled on tests requiring spatial orientation and sensori-motor coordination. Similarly, in an analysis of the ACE scores of Jewish and non-Jewish freshmen at the University of Pittsburgh, the Jewish boys were found to excel the non-Jewish on the linguistic part of the test (33). The Jewish boys also did relatively better on the linguistic than on the quantitative parts of the test, while the reverse was true of the non-Jewish boys. Surveys of American-born children of Italian immigrants have generally shown that the children do relatively well on performance tests and relatively poorly when examined with abstract or linguistic materials (6).

Such differences in intelligence test performance among various immigrant groups may, of course, be accounted for partly on the basis of the differential language handicaps discussed in an earlier section Cultural traditions, however, undoubtedly play a major part in producing these group differences in intellectual development. In Jewish families, there is a characteristically marked emphasis upon the formal aspects of education and upon "abstract" intelligence, to the almost total neglect of "mechanical" intelligence and manual dexterity. Italians, on the other hand, have a traditional and age-old admiration for manipulative arts and crafts. The skill exhibited in the production of a beautiful object, a complex object, or an object well adapted to its practical use is held in high esteem and encouraged from early childhood. Relatively little emphasis, however, is placed upon the more abstract types of talent.

The specificity of inter-group differences in behavior was recognized by Porteus on the basis of his extensive observation and testing of various racial groups in Hawaii, Australia, and Africa (56). For example, he reports that the Chinese groups which he surveyed excelled the Japanese in tests of the Binet type and in auditory memory span. But the Japanese excelled the Chinese in the Porteus Maze Tests and in all performance and mechanical aptitude tests. Similarly, Australian aboriginals scored relatively well on the Porteus Mazes, but fell below the African and Asiatic groups in any test depending upon speed. Although inclining toward a hereditarian interpretation of race differences, Porteus concludes: "Among the racial groups

mental development is not even. Advantages in certain tests are balanced very often by weaknesses in another" (56, p. 74).

Cultural Hierarchy of Behavior Functions. It might be suggested that racial or national groups could be arranged in a consistent hierarchy if we considered only the "higher mental processes." Tests of abstract abilities, for example, are usually considered to be more diagnostic of "intelligence" than those dealing with the manipulation of concrete objects or with the perception of spatial relationships. The aptitude for dealing with symbolical materials, especially of a verbal or numerical nature, is likewise regarded as the acme of intellectual attainment. The "primitive" man's skill in responding to very slight sensory cues, his talents in the construction of objects, or the powers of sustained attention and muscular control which he may display in his hunting behavior are regarded as interesting anthropological curios which have, however, little or no intellectual worth. As a result, such activities have not usually been incorporated in intelligence scales, but have been relegated to a relatively minor position in mental testing.

Upon closer analysis it will become apparent that this conception of intellect is itself culturally conditioned. By "higher mental processes" is usually meant those aspects or segments of behavior which are at a premium in our society. Intelligence tests would be very different if they had been constructed among American Indians or Australian aborigines rather than in American cities. The criterion employed in validating intelligence tests has nearly always been success in our social system. Scores on the test are correlated with school achievement or perhaps with some more general measure of success in our society. If such correlations are high, it is concluded that the test is a good measure of "intelligence." The age criterion is based on the same principle. If scores on a given test show a progressive increase with age, it may simply mean that the test is measuring those traits which our culture imparts to the individual. The older the subject, the more opportunity he will have had, in general, to acquire such aptitudes.

Thus it would seem that our intelligence tests measure only the ability to succeed in our particular culture. Each culture, partly through the physical conditions of its environment and partly through social tradition, "selects" certain activities as the most significant. These it encourages and stimulates; others it neglects or definitely

suppresses. The relative standing of different cultural groups in "intelligence" is a function of the traits included under the concept of intelligence, or, to state the same point differently, it is a function of the particular culture in which the test was constructed.

Since the current intelligence tests are a characteristic American development and since the testing of racial groups has been conducted largely by American psychologists, inter-group comparisons have generally been made with tests standardized within our culture. On such tests it is not surprising that most comparisons favor American subjects. What would happen if a test were constructed in a different culture by a procedure analogous to that followed in the preparation of our own tests? The instances in which this has been attempted are rare, but the results are enlightening.

In the course of an investigation by Klineberg (cf. 39) among the Dakota Indians, a "beadwork test" was devised in which a small sample of beadwork was shown to the subjects for four minutes; the sample was then removed and the subjects asked to reproduce it from memory on a loom. The test was applied to both white and Indian girls, all of whom were first taught how to do beadwork on a loom. As would be expected from their greater familiarity with this type of material, the Indian girls clearly surpassed the whites. Similarly, P. H. DuBois (21) standardized a Draw a Horse Test on Indian children, following closely the procedure of the Goodenough Draw-a-Man Test. In terms of age-grade placement and other criteria of "intelligence," the horse test proved to be more valid than the man test for these children. Moreover, when both tests were administered to white and Indian children, the whites excelled on the man-drawing and the Indians on the horse-drawing test. On the basis of the latter test, the 11-year old white boys tested in this study would have obtained an average "IQ" of 74!

Porteus (55) tried a similar experiment while working among the Australian aborigines. Having been impressed with the remarkable tracking skill of these people, he constructed a test with photographs of footprints, the task being to match the two prints made by the same foot. On this test, the Australians did practically as well as a group of 120 white high school students in Hawaii who were tested for comparison. In commenting upon these results, Porteus remarks:

> Allowing for their unfamiliarity with photographs we may say, then, that with test material with which they are familiar the aborigines' ability

CHAPTER

22

Racial versus Cultural Differences

IT IS APPARENT that the comparison of either the everyday achievements or the psychological test scores of different racial groups cannot in itself provide valid information regarding *race differences* in abilities or personality. We have seen that because of large group differences in many important aspects of the psychological environment, tests designed for one group may not have the same diagnostic significance when applied to another group. For example, an intelligence test standardized on white American urban school children is a valid measure of intelligence only for white American urban school children. Even if our only aim is to predict how well the individual will progress academically within the white American urban culture, this test could not adequately make such a prediction when given to a child reared under different conditions. An IQ of 60 on such a test would not have the same meaning or diagnostic significance when obtained by individuals with different environmental backgrounds. Thus if such an IQ were obtained by a white child of professional parents, it might indicate some structural deficiency which prevents normal intellectual development. But if the same IQ were obtained by a Negro child from a poor and isolated rural community, it might simply mean illiteracy. The prognosis for intellectual improvement in these two cases, given equal subsequent educational opportunities, would certainly be quite different.

To the extent that any two groups differ in both racial (biological) and cultural (environmental) factors, the mere comparison of their test performance yields ambiguous results. Such results not only fail to answer the theoretical question regarding the presence or absence of racial differences as such, but they are also of questionable value for the practical problem of diagnosis and prediction. Such predictions could be validly made only on the assumption that environ-

mental group differences are frozen or unchanging, a condition which is not at all characteristic of modern civilization.

In an effort to devise more fruitful experimental designs for the study of race differences in psychological functions, a number of investigators have combined *intra-racial* with inter-racial comparisons. Such an approach, although still falling short of a rigid control of conditions, permits a somewhat better analysis of group differences than is possible by a simple comparison of racial groups as a whole. Under this category may be included *psychological studies of hybrid groups,* in which the degree of race mixture is compared with psychological test scores. Investigations of *regional differences and migration* likewise permit comparisons among groups of the same race living under different environmental conditions. An especially promising procedure is represented by *cross-comparisons among racial and national groups*. Such comparisons can be made in many European nations, whose populations are composed of several subdivisions of the Caucasian race (cf. Ch. 20). The representatives of these racial sub-groups living within a single nation are, in general, exposed to relatively uniform cultural conditions. To a greater or less extent, they share the social traditions and customs of their country. Since all are members of the "white" race, social distinctions and discriminations are far less prevalent than is the case with Negroid or Mongoloid groups living among Caucasians.

Some investigations have been especially concerned with *racial* versus *cultural factors in the development of personality*. In such studies, cross-comparisons have also been made among racial and cultural groups, the latter groupings sometimes being represented by nationality and sometimes by more homogeneous and more narrowly defined cultural units. An interesting illustration of the relative role of cultural and biological factors in the development of characteristic group behavior is provided by studies on *gesture and cultural assimilation*. In the sections which follow, we shall consider typical investigations in each of the areas cited above.

PSYCHOLOGICAL STUDIES OF HYBRID GROUPS

It has been argued that if one racial group is "biologically superior" to another, a mixture of the two should produce individuals who are intermediate between the superior and inferior stocks. Moreover,

the greater the contribution of the superior race to the individual's heredity, the higher should be his intellectual level, according to such a view. It should be noted, however, that this problem, too, has its complications. In the first place, race mixture is often selective. This is especially true of those mixtures which are discouraged or frowned upon by society. In such cases, miscegenation may be confined largely to the socially and educationally inferior members of both groups. More often, perhaps, the selection occurs only in the group which is socially dominant, there being relatively little prejudice against such a mixture among members of the socially less favored group. This was doubtlessly the case when a "civilized" and a "primitive" group first came into contact. It has also been suggested (cf. 47) that a certain amount of selection may occur in the reverse direction, the superior individuals of the "lower" race being chosen more often for such unions. It is doubtful, however, whether such a selection is made on an intellectual basis to a significant degree.

A second important consideration is that the hybrid individual is usually more highly assimilated to the culture of the socially dominant group than is the full-blood. Because of the prevailing beliefs regarding the relative status of the two races, he is usually considered to be more capable than his full-blooded cousins. As a result, he is given better educational opportunities, admitted to more responsible positions, and afforded superior facilities for advancement in every way. A third and related point is the fact that mixed-bloods have been exposed more directly to the manners and customs of the dominant group than have the full-bloods. Whether the miscegenation occurs through legal marriages or illicit unions, the presence of the white parent will on the whole tend to bring about a closer contact with the white culture than is the case in families where no such mixture has occurred. As a result, the economic and the social level of the home, as well as the degree to which English is spoken at home, frequently differentiate hybrid from full-blood groups. This is particularly true of American Indian groups, which vary widely in their degree of assimilation to the white culture.

All these factors must be kept in mind in interpreting the findings of studies on hybrid groups. A few scattered investigations have been conducted on groups of mixed racial origin in Hawaii (46), Jamaica (8), Africa (14), and elsewhere. The most extensive data, however, have been collected on the American Indian and the American Negro,

owing to the relative accessibility of these groups in large numbers.

Hunter and Sommermier (26) administered the Otis Group Intelligence Test to 711 American Indians from a large number of different tribes, who were attending the Haskell Indian Institute at Lawrence, Kansas. The subjects ranged in age from 14 years upwards. Extent of race mixture was determined directly by an examination of ancestry records. No evidence of mixture with any race other than the white was found. The full-blood Indians formed the largest group, numbering 265; only 7 members of the entire tested sampling had less than ¼ Indian blood. As would be expected from the verbal nature of the test, the average Otis score of the Haskell group as a whole was much lower than the white norms, age by age. Analysis of performance on the separate parts of the test showed the Indians to be most inferior in the more highly verbal tests, such as analogies, opposites, matching proverbs, and narrative completion. With reference to race mixture, a correlation of .41 was found between total Otis score and degree of white blood, within the entire Indian group.

In a later study, Garth and his co-workers (20) analyzed the National Intelligence Test scores of 609 mixed-blood and 89 full-blood Indians attending Indian reservation schools in South Dakota, Oklahoma, New Mexico, and Colorado. A group of 67 white children was also tested for comparative purposes. The intelligence test scores again showed a steady rise with decrease of Indian blood, the averages for ¾-bloods, ½-bloods, and ¼-bloods being 74, 75, and

TABLE 48 *Correlation between Degree of White Blood and National Intelligence Test Scores within Each School Grade*

(From Garth et al., 20, p. 274)

Grade	*Number of Cases*	*Correlation*
Fourth	134	.70
Fifth	169	.76
Sixth	180	.22
Seventh	112	.23
Eighth	75	.24

77.5, respectively. The correlation between degree of white blood and test score proved to be +.42. The data were further analyzed in respect to separate school grades. In Table 48 will be found the num-

ber of cases in each grade as well as the correlation between National Intelligence Test score and degree of white blood within that grade. It should be noted that the distribution of white blood was similar in all grades and could not therefore account for the differences obtained.

These data suggest rather strongly an environmental interpretation of the correlation between degree of white blood and intelligence test performance. In the lower grades, those children with a larger percentage of white blood clearly excelled their fellows. In the three upper grades, however, the relationship is very low and barely significant. Thus continued education in a common school seems to reduce and even wipe out the apparent relationships with degree of Indian blood.

Klineberg (30) reports an absence of linear relation between degree of Indian blood and test performance in a group of 100 Yakima Indians in the state of Washington. The tests were taken from the Pintner-Paterson Performance Scale and were largely dependent upon speed. The Indians as a whole obtained lower scores than a group of 100 white boys who had been similarly tested. Comparison of full-blood and mixed-blood groups, however, gave conflicting results, the poorest scores having been obtained by those subjects with the most and those with the least Indian blood.

More recently, Telford (52) investigated specifically the effect of the cultural content of the test upon the performance of hybrids. The subjects were students at various Indian schools in North Dakota and Montana. The tests included two scholastic achievement tests, two group intelligence tests (Otis and Kuhlmann-Anderson), the Peterson Rational Learning Test (a non-verbal test), six performance tests from the Pintner-Paterson series, and the Goodenough Draw-a-Man Test. The results support the hypothesis that the superiority of mixed- over full-blood Indians, reported in some of the earlier investigations, is due to the greater familiarity of the mixed-bloods with English and with information based upon the white culture. In the achievement tests, Telford found the mixed-bloods superior to the full-bloods. A smaller but still significant difference in favor of the mixed-bloods was obtained on the intelligence tests. On the Peterson Rational Learning Test the two groups were equal, and on the performance and Goodenough tests the full-bloods showed a small, insignificant superiority.

Further evidence for the cultural hypothesis is provided by a study

conducted among the Osage Indians by Rohrer (49). Osage children were chosen for this study because they are the most nearly comparable to white children in socio-economic level, use of the English language, and schooling. Rapport and motivation in the testing situation are described as having been equally good among the Indian and white children tested. All the Indian children were attending either public schools or tuition schools, and were compared with white children attending the same schools. The proportion of Indian blood for each child, as determined by ancestry records, ranged from $\frac{1}{64}$ to 100%. On both the Goodenough Draw-a-Man test and the Otis test, no difference was found between groups differing in per cent of Indian blood. Moreover, the Indian children as a group did not differ significantly from the white controls, nor from the white norms, in *either* test. The number of cases tested, mean scores, and correlations with degree of Indian blood are given in Table 49.

TABLE 49 *Test Scores of Osage Indian Children in Relation to Degree of Indian Blood*

(Adapted from Rohrer, 49)

Test	*Number of Cases*		*Mean IQ*		*Correlation with Degree of Indian Blood*
	Indian	White	Indian	White	
Goodenough Draw-a-Man	125	125	103.80	102.92	.01
Otis S-A: Intermediate	110	110	100.05	98.05	.002

In investigations on the *American Negro*, ancestry records have not generally been available, so that degree of race mixture has had to be determined more indirectly on the basis of physical characteristics. In an early study on 907 Negro school children in three Virginia cities, Ferguson (13) reported a steady rise in psychological test performance with increasing proportion of white blood. Four simple psychological tests were administered: analogies, sentence completion, A-cancellation, and stylus maze. No anthropometric measures of racial characteristics were taken, the subjects being classified by inspection into four groups on the basis of skin color, hair color, and shape of head and face.

In a later study, Peterson and Lanier (44) administered a number of "ingenuity" tests as well as intelligence scales to 12-year-old Negro school children in Nashville, Chicago, and New York City. Several of the tests were non-verbal, an important consideration in the comparison of racial groups with diverse educational opportunities. Ratings of skin color on a 7-point scale were obtained on the Nashville and Chicago groups. In Table 50 are shown the correlations between lightness of skin and scores on the five tests employed.

TABLE 50 *Correlations betwen Lightness of Skin and Test Scores of 117 Negro School Children*

(From Peterson and Lanier, 44, p. 86)

Test	*Number of Cases*	*Correlation*
Binet Group Test	83	.18
Myers Mental Measure	75	.30
Rational Learning, Time Score	117	.05
Mental Maze, Time Score	113	.14
Disc Transfer, Time Score	119	.39

In view of the inadequacy of skin color as a criterion of race, more extensive measures were obtained on the group of 75 New York City subjects. Correlations were computed between score on the Yerkes Revision of the Binet Intelligence Scale and each of the four physical traits which were found to differentiate most clearly between white and Negro subjects. These correlations are shown in Table 51 below. As will be seen, the correlations are all too low to indicate a significant degree of relationship.

TABLE 51 *Correlations between Intelligence Test Scores and Anthropometric Measures on 75 Negro School Boys in New York City*

(From Peterson and Lanier, 44, p. 90)

Measure	*Correlation*
Nose width	−.11
Lip thickness	.07
Ear height	−.15
Interpupillary span	.01
Composite of these four traits	−.13

Klineberg (30), in the previously described investigation with the Pintner-Paterson Scale, also tested 139 Negro boys between the ages of 7 and 16 in rural sections of West Virginia. The correlations [1] between intelligence test score and three anthropometric measures indicative of degree of Negro blood are given below:

Nose width	−.083
Lip thickness	−.068
Black pigmentation	.12

As in the study of Peterson and Lanier, the relationship between test performance and index of Negro blood is negligible when objective anthropometric measures of race mixture are employed. In a group of 115 Negro men students at Howard University, Herskovits (24) likewise found no significant correlations between intelligence test scores and the same three anthropometric measures of Negroid characteristics.

Also relevant to the question of race mixture are the data collected on gifted Negro children. Witty and his students (27, 28, 29, 53, 54, 55) have reported a number of test surveys, case studies, and follow-ups of Negro children whose IQ's ranged from 120 to 200. In one such survey of published studies dealing with intellectually superior Negro children, Jenkins (29) assembled case records of 18 Negro children who tested above IQ 160 on the Stanford-Binet. It might be noted parenthetically that the results of all these studies are closely similar to those obtained by Terman and others on gifted white children (cf. Ch. 17). Intellectually superior Negro children, like white children of corresponding IQ, tend to excel in height, weight, and general physical development, they are on the whole superior in character and personality; and their parents have more than average education and tend to cluster in the higher occupational levels. For the present purpose, however, it is the racial background of such gifted Negro children that is of special concern.

In general, the distribution of white and Negro blood in such intellectually superior groups is no different from that in the general American Negro population. In one survey by Witty and Jenkins (54), 63 Negro school children with IQ's of 125 or higher were classified into four categories of race mixture on the basis of genealogical data secured from the parents. In Table 52 will be found the percentage of children falling into each of these categories for the

[1] The influence of age was ruled out by the partial correlation technique.

entire group of 63, as well as for a sub-group of 28 with IQ's of 140 or higher. For comparative purposes, the corresponding percentages for the general Negro population are also given. It will be seen that there is no consistent tendency for the proportion of white blood to be greater in the gifted groups than in the general Negro population. It is also interesting to note that the highest IQ in the group, 200, was obtained by a Negro girl whose ancestry showed no evidence of white mixture (53, 54).

TABLE 52 *Degree of White Blood among Negro School Children with IQ's of 125 or Higher*

(Adapted from Witty and Jenkins, 54, pp. 189-196)

Degree of White Mixture	*Per Cent in General Negro Population*	*Per Cent among Gifted Negro Children*	
		IQ 125 or Higher (N = 63)	*IQ 140 or Higher* (N = 28)
No white ancestry *	28.3	22.2	21.4
More Negro than white	31.7	46.1	42.8
About equal	25.2	15.9	21.4
More white than Negro	14.8	15.9	14.3

* Less than $\frac{1}{16}$th of mother's and of father's ancestry reported to be white.

Considerable caution should be exercised in generalizing from these findings, since the number of gifted Negro children included in such studies is quite small. Several attempts have been made to compute the percentage of Negro children falling within various segments of the distribution of intelligence, and to compare the resulting figures with similar figures for white children. Such a comparison is complicated, on the one hand, by the fact that the samplings employed may not be equally representative of the entire white and Negro populations, respectively, and on the other hand, by the fact that socio-economic and other environmental conditions are not equated in the two populations. It is not surprising, therefore, that estimates of the relative incidence of gifted children among Negroes and whites have varied so widely (cf., e.g., 19, 29). One fact which is clearly brought out, however, is that high intelligence is not *precluded* by any degree of Negro blood. Individual cases of highly gifted children can be found among Negroes of any degree of racial mixture or purity.

All in all, the available data on hybrid groups tend to uphold a cultural rather than a biological hypothesis of observed race differences in tested abilities. Among the previously cited findings in support of such a conclusion may be mentioned the following:

(1) The correlation between degree of white mixture and intelligence test score among Indian children tends to decrease as amount of education increases.

(2) The correlation between degree of white blood and test score among Indian children is highest for verbal tests and for tests depending upon information characteristic of the white culture. It is lower on non-language tests and drops to virtually zero when both speed and language are eliminated, as in the Goodenough test.

(3) Among Negroes, the correlations between test scores and degree of white mixture have in general been much lower than among Indians. Corresponding to this finding is the fact that differences in use of English and in assimilation of the white culture are much greater among Indians than among Negroes. These differences are quite closely related to degree of white mixture among various Indian groups. Moreover, mixed-blood Negroes are more likely to be classed socially with the Negro race, regardless of the amount of white blood. In the case of Indians, the mixed-bloods are more likely to be classified in accordance with their proportion of white blood, rather than being indiscriminately regarded as "Indian." This would make for more socially determined differentiation among Indians with different degrees of white blood than among Negroes with different degrees of white blood.

(4) In Negro studies in which classifications of race mixture were based upon objective physical measures, the correlations between test scores and degree of white blood were much lower than when inspectional criteria were employed. Skin color likewise tended to give higher correlations than other criteria which were equally good or better indices of white mixture. These findings suggest that it is not so much the actual amount of mixture as the general *observable* resemblance to the white race which was correlated with test performance. Such a general resemblance would play an important part in the

social contacts and everyday life opportunities of the mixed-bloods. In other words, it may have been the degree of *social acceptance*, rather than the amount of race mixture, that determined these correlations.

REGIONAL DIFFERENCES AND MIGRATION

Another procedure which may contribute to an analysis of the role of cultural and biological factors in group differences is the comparison of samplings of the same race living in different regions. This is especially fruitful when the regions present sharply differentiated environmental milieus. When feasible, the direct study of migrating groups before and after migration, or following different periods of residence in the new area, should permit a more clear-cut evaluation of contributing factors than is possible in static comparisons of regional groups.

It is a well-established fact that within any one racial group there are wide differences in tested abilities from one part of the United States to another. This was first vividly demonstrated by a state-by-state compilation of the army test results in World War I (1, 56). In some states, the median Alpha score of white enlisted men was as low as 41, in others as high as 79 or 80. To be sure, such regional differences may in part reflect differences in testing policy in the various areas, especially as regards the administration of Alpha or Beta (cf. Ch. 20). But it is unlikely that this factor could account wholly or even in large part for the differences obtained. Similarly, the proportion of foreign-born in each state could not have been a major factor, since most of the foreign-born were probably tested with Beta.[2] A number of interesting correspondences were found between the rank-order of the states in Alpha medians and their rank-order in certain environmental measures, such as socio-economic and educational indices. For example, the 41 states in which Alpha scores were available for at least 500 men were ranked for "educational efficiency." The latter was based on such records as percentage of daily school attendance, percentage of children attending high school, per capita expenditure for education, and teacher salaries. The

[2] In fact, a correlation of .61 was found between per cent of foreign-born population and Alpha median for each state. This correspondence probably resulted spuriously from the fact that the states with relatively large foreign populations were also the more highly industrialized and wealthier states, with better educational facilities.

two sets of ranks, for Alpha score and for "educational efficiency," correlated .72.

Similar geographical differences have been found in the AGCT scores obtained in World War II (9, 50). Marked variations in test performance were noted among the nine Service Commands, or major areas into which the country was divided in Selective Service classifications. In general, the southeastern and southwestern states had the highest rates of rejection for intellectual inadequacy, as well as the largest percentage of men in army grades IV and V on the AGCT (9). Even when men in the *same occupations* were compared, the samples from northern states generally obtained significantly higher AGCT scores than those from southern states (50).

These regional differences are just as characteristic of Negroes as they are of whites. In both World Wars, the army testing showed differences which were fully as large among the Negro samplings from different states or regions as among the white groups (9, 56). Moreover, the rank-order of the different areas was closely similar for both racial groups. Such results suggest that Negro test scores are as responsive to the environmental differences represented by the various regions as are white test scores. Of special interest in studies on the American Negro are comparisons between northern and southern Negroes. Differences between such groups undoubtedly result in part from the same socio-economic and educational differences which

TABLE 53 *Alpha and Beta Medians of Northern and Southern Negro Draft and of Native-Born White Draft*

(Adapted from Yerkes 56, p. 704)

Sampling	*Alpha*		*Beta*	
	N	Median	N	Median
White native-born	51,620	58.9	11,879	43.4
Northern Negro	2,850	38.6	1,737	32.5
Southern Negro	1,709	12.4	3,438	19.8

account for the differences between whites from northern and southern states. But they probably also reflect, to a certain extent, differences in the relative social position of the Negro in the North and the South.

In Table 53 will be found the Alpha and Beta medians of northern and southern Negro draftees in World War I, together with the num-

ber of cases in the samplings from which these medians were derived. The corresponding medians for the white native-born draft from the entire country are included for comparison. The Alpha and Beta results cannot be directly compared, because the scores are in different units and because the selective factors affecting Alpha and Beta samplings varied in different regions (cf. Ch. 20). The data do, however, serve to illustrate the large difference in average test performance between northern and southern Negroes.

This regional difference has been repeatedly corroborated in studies on Negro school children and college students. In the previously cited investigation by Peterson and Lanier (44), for example, large and highly significant differences in test scores were found between the New York City and the Nashville groups of 12-year-old Negro school children. In a more recent study by Roberts (48), the ACE scores of 253 Negro male college freshmen were analyzed. Comparisons were made in terms of parental occupation, veteran or non-veteran status, and northern or southern origin. Larger and more significant differences in ACE score were found between comparable groups from the North and the South than in any of the comparisons within a given geographical region. The regional difference persisted when comparisons were made between groups matched in occupational level of parents.

The latter study illustrates a finding which has been repeatedly demonstrated, viz., *formal education and socio-economic level are not enough* to account for differences in the tested abilities of northern and southern Negroes. When the amount of education is held constant, the regional differences are reduced but by no means eliminated (cf., e.g., Table 46, Ch. 21). The same point can be made regarding Negro-white differences in intelligence test scores. When only *amount* of education is held constant, part of the remaining group difference may be due to dissimilarities in the *quality* of education received by whites and Negroes, or by northern and southern Negroes.

There is some evidence, however, to indicate that even when Negro and white children in the same schools are compared, intelligence test performance favors the whites. This was demonstrated in an investigation by Tanser (51) on Canadian Negroes. Similarly, in a study conducted by Bruce (7) in a poor rural district in Virginia, the white children averaged higher than the Negroes on intelligence tests, even though an effort was made to control educational and

socio-economic variables. Superficially such findings appear to show a true "racial," or biological, difference which persists even when inequalities of education and socio-economic level are eliminated—and they have been given this interpretation by some writers. On the other hand, it should be noted that no study has satisfactorily controlled *both* socio-economic level and educational facilities in Negro-white comparisons. Thus in Tanser's study, although the Negro and white children attended the same schools, there were significant differences in socio-economic level between the two groups. In the investigation by Bruce, education was equated only by choosing Negro and white schools which had the same teacher-pupil ratio, other probable differences between these schools remaining uncontrolled. In the same study, sub-groups of 49 Negro and 49 white children were matched in Sims socio-economic ratings; but the author admitted that this scale was unsuited to the groups studied because it does not discriminate adequately at the lower socio-economic levels, where most of the subjects fell.

The psychological environment, moreover, includes much more than formal schooling and socio-economic class. The many subtle emotional and motivational influences associated with minority group status and with traditional stereotypes still remain as uncontrolled factors in all these group comparisons. It is interesting to note, for example, that in Tanser's study the white children *attended school* much more regularly than the Negro (51). Within the entire sampling of white children tested, school attendance averaged 93.38%; within the Negro group, it averaged 84.77%. Such factors as family traditions, social expectancy, and outlook for adult opportunities may all be reflected in these school attendance figures. Group differences which are the result of a large number of variables obviously cannot be wiped out by holding one or two variables constant.

One of the frequently cited results of the army testing in World War I was that the median Alpha scores of Negroes from Illinois, New York, Ohio, and Pennsylvania were somewhat *higher* than the median Alpha scores of whites from Arkansas, Georgia, Kentucky, and Mississippi. More than twenty five years after these particular data were collected, they were resuscitated, re-analyzed, and made the subject of a scientific storm in a teakettle (cf., e.g., 2, 3, 5, 15, 16, 17, 18, 19). Obviously these data were not meant to provide an adequate comparison of Negro and white performance, since the

samplings were hardly comparable. The groups were not even representative of all Negro or white recruits from the respective states, since the proportion of men tested with Alpha and Beta differed from state to state and between Negroes and whites. The relative standing of these groups, of course, follows from the fact that there were large regional differences among *both* Negroes and whites. The data do provide a vivid illustration of the extent of *overlapping* between the white and Negro distributions. Not only could many *individuals* be found in the lower (Negro) distribution who excelled individuals in the higher (white) distribution, but also local *groups* could be found in the lower-scoring population which excelled other local groups in the higher-scoring population.

In explanation of regional differences in intelligence test performance, two contrasting hypotheses have been proposed: one in terms of *environmental handicap*, the other in terms of *selective migration*. The former attributes the regional differences to inequalities in home conditions, educational facilities, and other opportunities for advancement. The latter proposes that the more intelligent and progressive individuals, who have more initiative and are better able to adjust to new surroundings, are more likely to migrate to the more desirable areas. The one hypothesis maintains that superior ability is a *result* of migration to a more favored area, the other that the migrating individuals were superior to begin with. Although the selective migration hypothesis is commonly coupled with a hereditary interpretation of regional differences in ability, it should be noted that this does not necessarily follow. Thus it is likely that persons of superior educational and socio-economic level are more often aware of the opportunities offered by migration to a better area. This would be true regardless of whether hereditary or environmental factors were initially responsible for the higher educational and socio-economic status of such individuals. These persons may in turn have more intelligent offspring, not necessarily because of better "genetic stock," but because they provide their children with a more stimulating home environment. Logically, therefore, the selective migration hypothesis is equally consistent with a predominantly hereditary or a predominantly environmental determination of individual differences. If, on the other hand, regional differences can be shown to have developed *after* migration, then they can be explained only in environmental terms.

A number of investigations have been specifically designed to test the selective migration hypothesis with reference to northern and southern Negroes (32, 38, 41). In one survey on several thousand Negro school children in Washington, D. C., Long (38) found significant differences in mean IQ in favor of Washington-born children. In separate comparisons made among first, third, and fifth grade children, the critical ratios of these mean IQ differences (diff/σ_{diff}) ranged from 3.48 to 6.71. In other words, the children who had migrated to Washington—in most cases from inferior southern communities—were not as intelligent as those born and reared in Washington. In a more intensive study of a random sample of the migrant children, a significant positive correlation was found between length of residence in Washington and IQ.[3]

The most ambitious effort to check the applicability of the hypotheses of selective migration and of environmental handicap is to be found in the series of investigations by Klineberg and his students (32). The problem was approached in two ways. First, the relative intellectual status of Negro children whose families had migrated to the North was investigated by comparing their former grades in southern Negro schools with the norms for those schools. In this part of the study, the records of 562 Negro children who had moved to the North from three southern cities[4] were examined. Since all grades were transmuted into a percentile scale, a score of 50 represents the average status, and this figure may be employed as a standard of comparison. The average percentile rating of those children who had moved to the North proved to be 49.3, which is not significantly different from the general average. It is thus apparent that, at least in these groups, there was no tendency for the initially superior children to migrate.

A second approach to the problem involved the comparison of intelligence test scores obtained by groups of Negro school children who had lived in New York City for different periods of time. The subjects were examined with a variety of standard tests, including the Stanford-Binet, performance scales, and several common group tests. Over 3000 10- to 12-year-old Negro children in the Harlem district of New York City were tested. The subjects in the different

[3] The correlation ratio (*eta*) was computed, since the relation between length of residence and IQ was found to be curvilinear.

[4] Nashville, Tenn., Birmingham, Ala., and Charleston, S. C.

residence groups were equated for age and sex; they attended the same schools and were approximately equal in socio-economic background, the only important difference between them being the number of years spent in New York City. A group of Negro school children born in New York City was also included for comparison. Special checks were employed to demonstrate that the differences between the various residence groups could not be attributed to difference in the proportion of white mixture, nor to a progressive decline in the quality of migrants coming to New York in successive years.

TABLE 54 *Relation between Length of Residence in New York City and Intelligence Test Scores of Negro School Children*

(Adapted from Klineberg, 32)

National Intelligence Test		
Years of Residence	Number of Cases	Mean Score
1–2	150	72
3–4	125	76
5–6	136	84
7–8	112	90
Over 8	157	94
Northern-born	1017	92

Stanford-Binet		
Years of Residence	Number of Cases	Mean IQ
Less than 1	42	81.4
1–2	40	84.2
2–3	40	84.5
3–4	46	85.5*
Over 4	47	87.4
New York-born	99	87.3

Minnesota Paper Form Board		
Years of Residence	Number of Cases	Median Score
1–2	27	39.00
3–4	25	26.67
5–6	30	31.88
7–8	23	37.50
9–10	25	37.50
Over 10	41	37.50
New York-born	223	41.61

Pintner-Paterson		
Years of Residence	Number of Cases	Mean Point Score
Less than 2	20	142.5
2–5	20	139.8
Over 5	20	152.1
Northern-born	50	164.5

* This figure is misprinted as 88.5 in the Klineberg monograph (32, p. 46).

In Table 54 will be found the mean scores of each residence group on the National Intelligence Test, the Stanford-Binet, the Minnesota Paper Form Board, and an abbreviated form of the Pintner-Paterson

Performance Scale In both the National Intelligence Test and the Stanford-Binet, there is a progressive rise in average score with increasing length of residence in New York City The means and SD's obtained with the National Intelligence Test, which was given to the largest number of cases, are shown graphically in Figure 94 It is interesting

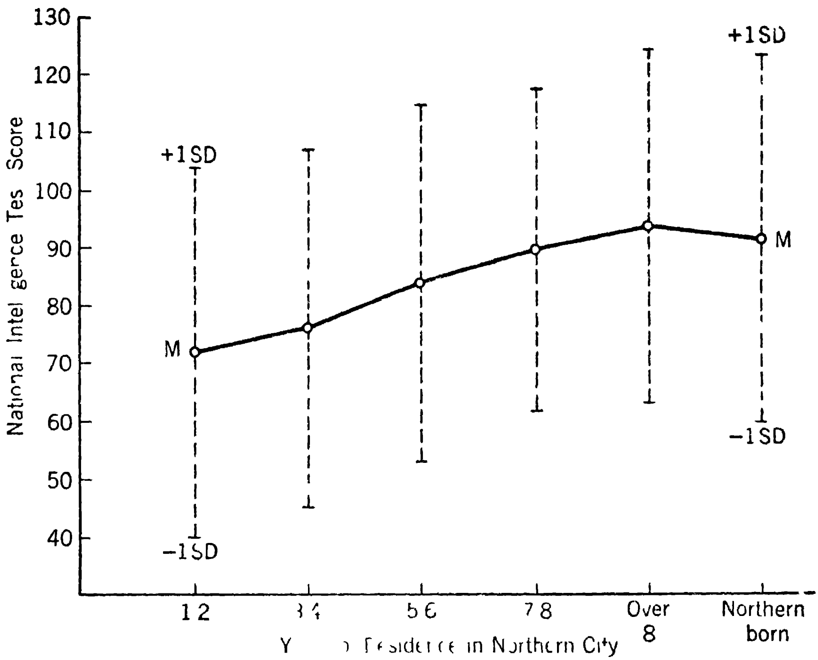


Fig. 94. Mean and SD of National Intelligence Test Scores of Negro Children in Relation to Length of Residence in a Northern City (Data from Klineberg 32 pp 26 31 [illegible])

to note that the groups born in the North or in New York City are not superior to those who were born in the South but had lived in New York for a long period This fact alone with the consistent increase in score with length of New York City residence, tends to support the environmental rather than the selective migration hypothesis Although overlap is large and differences between adjacent groups are too small to be statistically significant the more extreme differences in National Intelligence Test or Stanford-Binet means are significant at a high level of confidence We should hardly expect differences of only one

or two years in length of residence to affect test performance. Longer periods in a more favorable environment are, however, clearly reflected in these rising means.

The contrast between the trend observed in the verbal tests, on the one hand, and in the non-verbal tests, on the other, is also of interest in this connection. The rise in test score with increasing length of New York City residence is more consistent in the case of the National Intelligence Test and the Stanford-Binet than in the case of the Pintner-Paterson and the Minnesota Paper Form Board. This is to be expected, since the two verbal tests are much more highly dependent upon the type of information which would favor children reared in the urban New York environment. With regard to the two non-verbal tests, the Pintner-Paterson and the Minnesota Paper Form Board, the latter is even less dependent upon information of a specific cultural nature. And the results did show a more marked trend in the Pintner-Paterson than in the Minnesota Paper Form Board. All differences on the Minnesota test were insignificant, while a few of those found on the Pintner-Paterson approached the commonly accepted levels of significance. Thus, on the whole, the available evidence concerning regional differences among Negro school children favors the environmental hypothesis more strongly than it does the selective migration hypothesis.

CROSS-COMPARISONS AMONG RACIAL AND NATIONAL GROUPS

Cross-comparisons among individuals classified into racial and into national categories have been made with immigrant groups in this country, as well as with the parent populations in Europe. The first extensive effort to compare the intelligence test scores of European immigrant groups in America was Brigham's analysis of the Army data obtained during World War I (6). Brigham computed the mean combined scale scores[5] for 12,492 foreign-born draftees, classified according to country of birth. The resulting hierarchy of national groups, however, was of little significance in itself because of the uncontrolled operation of many of the factors discussed in Chapters

[5] The "combined scale" was a means of transmuting scores on Alpha, Beta, and individual tests into comparable units, in order to permit the direct comparison of individuals who had taken different tests.

20 and 21. For example, the two highest means were obtained by groups from English-speaking countries. Similarly, a consistent tendency was found for the mean combined scale score to rise with increasing length of residence in America, regardless of nationality.

Brigham further undertook to compare Nordic, Alpine, and Mediterranean sub-groups within the same sampling of foreign-born soldiers. For this purpose, he employed rough, available estimates of the proportion of Nordic, Alpine, and Mediterranean elements in each country. France, for example, was estimated as 30% Nordic, 55% Alpine, and 15% Mediterranean; Sweden, as 100% Nordic; Roumania, 100% Alpine; Germany, 40% Nordic and 60% Alpine. The distributions of intelligence test scores for each national group were then cut according to these proportions and recombined into Nordic, Alpine, and Mediterranean distributions. For example, all the Swedish scores were classified under Nordic, all the Roumanian under Alpine. In those cases in which more than one racial group was represented within a single nation, the average score of that national group was allotted proportionately to each racial group. Thus 40% of the German sampling was entered under Nordic and 60% under Alpine. Since there was no way of determining in which portion of the national distribution of scores the Alpine and Nordic individuals fell, all subjects were given the *average score* of their respective national group. By this method, Brigham found the Nordics to have a significantly higher mean score than the Alpines, and the Alpines to have a significantly higher mean than the Mediterraneans.

It is apparent that this procedure involves a logical fallacy in so far as it assumes the absence of differences in score between racial groups within a single nationality, and at the same time it undertakes to prove the existence of just such a difference among racial groups. Since no differentiation was made among individual members of different races within any single national group, the subjects being chosen indiscriminately from the entire distribution of national scores, nothing was really gained by the reclassification into Nordic, Alpine, and Mediterranean. Moreover, the same uncontrolled factors which rendered the comparison of national averages invalid in this study also operated in the comparison of the three racial groups. For example, the comparison of English-speaking and non-English-speaking Nordic groups revealed a significant difference in favor of the former.

An attempt to classify individuals more empirically into racial groups was made by Hirsch (25) in an investigation of children of immigrants in the United States. The main group of subjects consisted of 4983 Massachusetts public school children ranging in age from 5½ to 18 and in school grade from the first to the ninth. In social and occupational level the group was quite homogeneous, all the subjects living in small manufacturing communities. There was no segregation of national groups into districts and all the children attended the same schools. Group intelligence tests were chosen which relied somewhat less on language and on speed than is usually the case, although these factors were by no means eliminated.[6] The children were first classified into national groups on the basis of parents' birthplace. An "American" group of native parentage was also included for comparative purposes.

TABLE 55 *Mean IQ's of American School Children of Foreign Parentage*

(From Hirsch, 25, p. 287)

Nationality	*Number of Cases*	*Mean IQ*	*SD*
Polish Jews	75	102.8	14.55
Swedish	232	102.1	15.48
English	213	100.7	14.85
Russian Jews	627	99.5	14.58
Germans	190	98.5	15.09
Americans	1030	98.3	15.87
Lithuanians	468	97.4	13.89
Irish	214	95.9	16.08
British Canadians	155	93.8	14.67
Russians	90	90.9	12.93
Poles	227	89.6	12.96
Greeks	270	87.8	15.12
Italians	350	85.8	11.94
French Canadians	243	85.3	14.55
Portuguese	671	82.7	13.47

The results of this analysis are shown in Table 55. Most of the differences between the average IQ's of these national groups were

[6] The following tests were administered in different grade levels:
Pintner-Cunningham Primary Scale—first grade
Dearborn Test A—second and third grades
Dearborn Test C—fourth grade upwards
To reduce the role of speed, all tests were given with a longer time limit than is specified in the standardized directions.

statistically significant. The same rank-order of nationalities was obtained when the groups were compared in the percentage of "very superior intelligence" and of "borderline deficiency." The relative status of the national groups also agreed in general with that reported in previous investigations on children of foreign parentage. Taking the national groups as a whole, Hirsch found no evidence for the consistent superiority of any one racial group. Thus among the eight highest entries in Table 55 are to be found two predominantly Nordic groups (English and Swedish), two which are largely Alpine (Germans and Lithuanians), one predominantly Mediterranean (Irish), and three composite or mixed groups (Americans, and Polish and Russian Jews).

In order to arrive at a somewhat more accurate determination of "race," Hirsch classified each individual into a racial type on the basis of eye and hair color. All subjects, irrespective of their national descent, were divided into three major categories: the "*blond type*" with light hair and blue, gray, or hazel eyes; the "*brunette type*" with black hair and gray, hazel, brown, or black eyes; and the "*mixed type*" exhibiting all other combinations of hair and eye color. The blond type was taken to correspond roughly to the Nordic and the brunette to the Mediterranean race. The mixed type would of course include Alpines as well as mixtures of any of the three racial stocks. This method of classification is, to be sure, crude. Hair and eye color are not generally considered to be very valid criteria of race. The analysis is, however, suggestive as a first attempt in the direct classification of individuals into racial categories.

The results of this analysis likewise lent no support to a racial interpretation of group differences in intelligence test scores. No one of the three physical types was *consistently* superior or inferior within all national groups. Thus among the representatives of one nation the blonds stood first; among those of another nation the brunettes led. The differences *between* physical types, furthermore, were much smaller than those *within* a single type. The differences in IQ between any two types within a single nation ranged from 0.1 (between blond and mixed-type French Canadians) to 6.7 (between brunette and mixed-type Poles). The differences between the lowest and highest national averages within any one physical type, on the other hand, were all considerably larger. Thus the mean difference in IQ between the highest and lowest blond groups was 14.8; between the highest

and lowest brunette groups, 18.1; and between the highest and lowest mixed groups, 21.3. These cross-comparisons between national and physical or "racial" categories thus suggest that the obtained differences are more closely linked with national than with racial background.

The chief weaknesses in Hirsch's study are: (1) the likelihood that the samplings tested were not representative of their national populations because of selective factors in immigration; and (2) the use of very crude criteria for racial classification. Both of these limitations were avoided in a study conducted by Klineberg (31) in Europe, in which an attempt was made to obtain as pure samples of Nordics, Alpines, and Mediterraneans as possible. The subjects were 700 10- to 12-year-old school boys in rural sections of France, Germany, and Italy.[7] The samples were taken from those geographical areas in which ethnic maps showed a predominance of pure strains of each of these three racial groups. Only children who had themselves been born in the particular area, and both of whose parents had likewise been born in the same area, were included in the study. The subjects were further selected on the basis of three physical

TABLE 56 *Comparison of National and Racial Groups on a Performance Scale*

(From Klineberg, 31, p. 27)

Group	*Province*	*Number of Villages Covered*	*Performance Scale Score* Mean	Median	Range
1. German Nordic	Hanover	17	198.2	197.6	69–289
2. French Mediterranean	Eastern Pyrenees	12	197.4	204.4	71–271
3. German Alpine	Baden	10	193.6	199.0	80–211
4. Italian Alpine	Piedmont	10	188.8	186.3	69–306
5. French Alpine	Auvergne and Velay	19	180.2	185.3	72–296
6. French Nordic	Flanders	13	178.8	183.3	63–314
7. Italian Mediterranean	Sicily	9	173.0	172.7	69–308

[7] Rural groups were chosen since too much intermixture had occurred in urban districts to yield a sufficient number of "pure types." Three city groups, in Hamburg, Paris, and Rome, were also tested for comparative purposes. The results of this testing will be reported in the following chapter.

criteria: eye color, hair color, and cephalic index. No subject was retained unless he fell within the specified limits for his racial group in all three criteria. The groups were comparable in socio-economic and occupational levels, the differences among them in these respects being relatively slight.

Each subject was examined individually with an abbreviated form of the Pintner-Paterson Performance Scale, consisting of six tests.[8] Brief oral instructions were given in the subject's native language. Performance was scored in terms of speed as well as accuracy. In Table 56 will be found the mean, median, and range of scores within each group. The geographical location of the group and the number of villages covered are also given. The number of cases is exactly 100 in each of the seven groups.

The scores show marked variations among different samples of the same racial group. The alleged Nordic-Alpine-Mediterranean hierarchy is not maintained. Although the highest mean score is obtained by a Nordic group, the highest median is found in a Mediterranean group. Similarly, the rank-order of the racial groups *within* any one nation is inconsistent. Thus in France the Mediterranean group is best, the Alpine intermediate, and the Nordic poorest; whereas in Germany the Nordic is superior to the Alpine sampling, and in Italy the Alpine is superior to the Mediterranean. The marked *overlapping* of groups, as indicated by the range, should also be noted. When all Nordics, Alpines, and Mediterraneans are compared, regardless of nationality, the following mean scores are obtained:

Nordic	188.5
Alpine	187.5
Mediterranean	185.2

None of these differences is statistically significant. The variations from one Nordic sample to another, on the other hand, are large and significant. The same is true of the other two racial groups. Thus there is a difference of 24.4 points between French and Italian Mediterraneans; one of 19.4 points between German and French Nordics; and one of 13.4 between German and French Alpines.

The tests employed in this investigation are, of course, quite limited in the type of function which they measure. Moreover, within the age

[8] The Knox Cube, and the Triangle, Healy A, Two-Figure, Five Figure, and Casuist form boards.

range covered by the study, individual differences in score may reflect largely differences in speed of work, since the tasks are relatively easy for older children and adolescents. A repetition of this study with improved measuring instruments made possible by current developments in psychological testing would be a valuable addition to our understanding of group differences.

Within the behavior sampled by the tests which were employed, Klineberg's results did clearly demonstrate that the obtained differences among national groups could not be attributed to the "racial" composition or to the proportion of Nordics, Alpines, and Mediterraneans in each country. Because of the variations found among different samples of the same nation, Klineberg proposed that the differences may not even be national in scope, but should be envisaged in terms of smaller cultural units. That the differences are the result of environmental rather than hereditary factors is suggested by two considerations. In the first place, the predominance of a single inbred family strain in any one of the samplings tested is very unlikely because of the wide area covered. It will be recalled that from 9 to 19 villages were canvassed for each single sampling. In the second place, very interesting parallelisms were found between the cultural, economic, and educational conditions in any one region and the intelligence test performance of its inhabitants.

Although not concerned with national groupings, a re-analysis by Mann (40) of data collected by Porteus (45) may be included at this point, since this analysis is likewise based upon cross-comparisons among the same individuals classified with respect to biological and cultural criteria. Following a series of investigations on several native groups in Australia and South Africa, Porteus had concluded that the Australian aboriginals were racially superior to the Africans in the functions measured by the Porteus Maze Tests. He argued against an environmental explanation of the obtained differences, on the grounds that the environment of the Australian groups was actually more "repressive" than that of the Africans. This assertion he based principally on the greater scarcity of food and water in the habitat of the Australian groups. In itself, such an inference is questionable, since several of the African groups studied had to "contend not only with some of the most dangerous wild animals on earth but also with some of the fiercest native tribes," while the Australians were unmolested by either.

TABLE 57 *Porteus Maze Test Performance by Native Peoples of Australia and South Africa*

(Adapted from Porteus, 45, p. 257, and Mann, 40, p. 389)

Tribal Groups Classified According to Race			*Tribal Groups Classified According to Schooling Facilities*		
Race	N	Mean MA on Maze Tests	Schooling Facilities	N	Mean MA on Maze Tests
Australian	128	10.89	Mission or government	208	11.26
African	207	10.27	None	127	9.27
diff./σ_{diff}* = 1.8			diff./σ_{diff} = 6.0		

* In computing the σ_{diff}, Mann used an approximation of the SD's owing to the fact that he did not have access to the original scores. Since the contrast between the two types of comparison is so striking, however, it is unlikely that the computation of the precise SD's in each case would have affected the conclusion.

Porteus' data were subsequently reclassified by Mann with respect to schooling opportunities. Some of the tribes tested had had access to mission or government schools, while others had not. When these tribes were grouped, first, in terms of racial category (African or Australian), and secondly, on the basis of schooling facilities, the results shown in Table 57 were obtained. It is apparent that the former classification yields a small and rather insignificant difference, while the latter gives a much larger difference which is significant at a high level of confidence. Test scores in this study were thus more closely related to schooling than to racial category.

RACIAL *versus* CULTURAL FACTORS IN THE DEVELOPMENT OF PERSONALITY

Popular opinion has consistently attributed characteristic temperamental qualities to each race or nationality. Group differences in personality are held to be even greater than in ability, and the belief in such emotional differences persists even when intellectual equality is granted. Such familiar stereotypes as the Irish wit, the excitability of the South European or "Latin" groups, the easy-going nature of the American Negro, the stolidity of the American Indian, the composure of the Englishman, and a host of similar characterizations have become a part of our daily vocabulary.

In a number of investigations, paper-and-pencil personality tests have been administered to samplings of various groups living in this country, including European immigrants, Negroes, American Indians, and Orientals. Some investigators report no significant differences among the racial or national groups compared. Others have found slight differences, usually in the direction expected from tradition and popular belief. On the whole, the results of these studies are very difficult to interpret, partly because of the dubious validity of many of the tests, and partly because of the unrepresentative nature of some of the samplings employed. The comparison of Negro and white college students, for example, would be subject to a differential operation of selective factors in the two groups (cf. Ch. 20).

Moreover, on personality tests, even more than on tests of intellectual functions, a given test item may have a different meaning for Negroes and whites—or for any groups with very dissimilar experiential backgrounds. Even if all specific terms in the item are interpreted in an identical manner and with reference to the same standard by Negroes and whites, the same response may have a different diagnostic or prognostic significance when given by a Negro and by a white subject. Thus the statement that one is being discriminated against by many of his associates might indicate undue suspiciousness or even paranoid tendencies in a white respondent, but it might indicate only a realistic awareness of social attitudes in a Negro respondent.

An investigation conducted by Klineberg, Fjeld, and Foley (34) is of special interest, since it represents another application of the technique of *cross-comparisons* among cultural and biological groupings. Over 400 male and female students attending eight different institutions of collegiate rank in New York City and its environs were examined with a series of personality tests. The tests included the Bernreuter Personality Inventory, the Allport-Vernon Study of Values, an honesty test (Maller Test of Sports and Hobbies), and two tests specially devised for use in this investigation, one to measure suggestibility and the other persistence.[9] The subjects were classified into Nordic, Alpine, and Mediterranean groups on the basis of cephalic index, eye color, hair color, and skin color.

The mean scores of Nordic, Alpine, and Mediterranean groups on each test are given in Table 58, the data for the two sexes

[9] For a fuller discussion of these tests, the reader is referred to Chapter 13, in which another part of the same investigation was reported.

TABLE 58 *Mean Scores of Nordics, Alpines, and Mediterraneans on Personality Tests*

(From Klineberg, Fjeld, and Foley, 34)

Test	*Male*			*Female*		
	Nordic (N = 47)	Alpine (N = 49)	Mediterranean (N = 54)	Nordic (N = 64)	Alpine (N = 43)	Mediterranean (N = 45)
Allport-Vernon Study of Values:						
Theoretical	29.20	32.19	32.74	28.84	29.14	28.59
Economic	27.80	29.32	28.87	27.48	26.66	25.50
Esthetic	27.95	29.35	27.34	33.27	32.14	37.39
Social	31.32	34.75	31.89	31.98	32.45	31.13
Political	29.96	30.33	30.87	30.29	30.24	30.54
Religious	33.86	24.07	28.35	28.08	29.36	26.84
Bernreuter:						
B1-N: Neuroticism	−54.11	−35.37	−56.17	−52.44	−44.79	−20.29
B2-S: Self-sufficiency	43.13	37.55	29.59	14.48	13.91	2.62
B3-I: Introversion	−25.11	−14.76	−27.13	28.95	−22.60	−7.91
B4-D: Dominance	46.72	36.73	48.91	46.05	31.91	23.91
F1-C: Self-confidence	−28.17	−2.14	−24.43	15.67	−7.37	23.16
F2-S: Sociability	7.85	−2.76	−18.13	19.31	−16.44	−21.78
Suggestibility	11.77	11.67	11.59	12.53	10.93	11.04
Persistence *	8.13	11.37	8.82	9.40	9.60	8.39
Honesty	95.48	96.84	96.70	99.40	97.94	99.00

* Not all subjects were given these tests

being reported separately. On the Allport-Vernon Study of Values, only one of the differences is significant in each of the sex groups. Among the women there is a significant difference between Mediterraneans and Alpines in the mean score for "æsthetic value." This difference is 3.26 times as large as its standard error, the higher mean occurring in the Mediterranean group. Among the male students a significant difference was found between Nordic and Alpine groups in the mean score for "religious value." This difference was in favor of the Nordics, the critical ratio being 4.64 Upon further analysis, both of these differences seemed to be rather closely linked with institutional groupings. Thus the highest mean score for "religious value" was obtained in a Catholic college for men in which were found only three Alpines. This would tend to pull down the mean of the Alpines in relation to those of the other two racial groups. Similarly, among the female subjects, the highest scores in "æsthetic value" were obtained in an institution which encourages the æsthetic attitude, as is evidenced by a large and popular art department. This institution furnished a relatively large number of Mediterraneans, thus raising the mean "æsthetic value" score of the latter group. None of the other Allport-Vernon scores yielded significant differences between racial groups.

None of the differences in the six Bernreuter scores proved to be statistically significant in either male or female group. Likewise, in the three remaining tests, i.e.. suggestibility, honesty, and persistence, no significant group differences were found.

It is apparent that in the personality traits measured in this study the differences among Nordics, Alpines, and Mediterraneans within college samplings are very slight. Nor can it be argued that the lack of differentiation among these groups was due to the homogeneity of college students in the characteristics under investigation. Although relatively homogeneous in intellectual traits, college students exhibit large individual differences in personality development. This is borne out by the very wide ranges and SD's found within each group. It may also be mentioned that, as a result of the wide range covered by each group, a large and almost complete overlapping of the distributions of Nordics, Alpines, and Mediterraneans was obtained on each test.

In sharp contrast to the predominantly small and insignificant differences found between racial groups, many large and significant differences in mean score were obtained among the academic institutions

covered by this investigation. Several of these differences were many times larger than would be required to meet the usual standards of statistical significance. In both male and female samplings, the Allport-Vernon scores showed the largest differences. These differences agreed closely with well-known characteristics of the institutions under consideration. Thus in "religious value," the one Catholic college in the group obtained the highest mean score; the lowest mean was found in an institution whose student body was traditionally radical, agnostic, and of relatively low socio-economic level. The difference between these two means was 14.49 times as large as its standard error. It is interesting to note that another very large difference was obtained between the same two institutional groups in "theoretical value." In this case, however, the difference was in favor of the latter group, the critical ratio being 7.18. On the Bernreuter scales the differences were not so marked, although many were statistically significant. The tests of suggestibility, honesty, and persistence yielded relatively small and insignificant differences.

Whatever the cause of these institutional differences, it cannot be "race" in the biological sense, since the differences disappear when individuals are classified according to the physical criteria of race. The explanation of these personality differences from one institution to another is not difficult to find. In the first place, *selection* obviously operates in the students' enrollment in any particular institution. Individuals with certain attitudes and emotional characteristics will be more readily attracted to those institutions which are by tradition congenial to such traits. The evidence indicates, however, that such selection operates on the basis of the economic and cultural group in which the individual was reared rather than in terms of race. In the second place, *attendance in a particular institution* will itself foster the development of certain personality traits through the resulting social contacts and other direct stimulating circumstances.

In recent years, much has been said and written about *"national character."* The broad geographical scope of World War II brought about a sudden realization of the need for more knowledge regarding the customs, attitudes, and other psychological characteristics of many different cultures, including both allies and enemy nations. If we clearly recognize that "national character" is a cultural rather than a racial concept, the study of such national differences will not only yield results of practical value, but may also contribute to a better

understanding of the nature and causes of such group differences. Many of the techniques available for such comparative studies of different cultures, as well as some of the pitfalls to avoid, have been summarized by Klineberg (33).

Some psychologists and anthropologists consider the Rorschach test to be a promising instrument in this field of research, because of its relative independence of language and other culturally restricted content (23). Preliminary results have been reported on a number of American Indian groups and other cultures in which questionnaire methods would be quite unsuitable. The Rorschach test was also included by C. DuBois (10) in her intensive field study of the people of Alor, an island in the Netherlands East Indies. It must be remembered that the validity of many of the proposed diagnostic interpretations of specific Rorschach responses has not yet been satisfactorily established, even within any particular segment of our own culture. To what extent the various response characteristics may have the same significance in different cultures is also a matter that requires further study.

Among other available techniques cited by Klineberg (33) are those utilizing "laboratory" or performance-type tests, in which the subject's response to such situations as failure or frustration is observed. Considerable care must be exercised in generalizing from such a test, however, since the specific tasks may vary in importance for individuals of different cultures and thus motivation may not be comparable. Another possible source of data is to be found in the many descriptive accounts of "national character" which have appeared, including both the more journalistic, popular reports and the more technical surveys by anthropologists and sociologists.[10] The analysis of cultural products —such as humor, drama, moving pictures, literature, and popular songs—has also been a favorite approach. Even the examination of

[10] Cf., e.g., Benedict (4), Gorer (22), Lynd (39), Mead (42). For other references, cf. Gillin (21).

In 1948, an extensive project on national character was begun by a group of Columbia University anthropologists, under the auspices of the Psychological Branch of the Medical Sciences Division of the Office of Naval Research (cf. 36). This project, which was originally under the direction of Dr. Ruth Benedict, involves the application of anthropological methods to the study of a number of contemporary literate cultures. Studies on immigrant groups in New York City are being supplemented by field studies in the countries concerned. Psychological tests, interviews with representatives of different groups, and the analysis of cultural products are among the techniques being employed.

existing stereotypes may be helpful, if the stereotypes are recognized as such and are used only as leads for further analyses. Certain vital statistics, such as the frequency of psychoses or crimes of various types, may likewise provide useful data if supplemented by other information. For example, the relative frequency of homicide in defense of family honor in one culture, or of suicide in defense of individual honor in another, may furnish fruitful clues to the understanding of other characteristic behavior.

A method developed principally by anthropologists for use in relatively simple cultures, but subsequently extended to the analysis of national differences, is based upon a study of the child-training practices followed by different peoples (cf., e.g., 21, 22, 37). Feeding schedules, methods of toilet training, disciplinary techniques, and other child-rearing procedures are compared among different cultures. Some investigators have claimed that the characteristic adult attitudes in any one culture may depend in part upon the degree to which such childhood experiences were characterized by austerity, rigidity, aloofness, informality, emotional warmth, and the like. The available evidence for such claims, however, is extremely meager and of dubious significance (cf. 43).

All these methods represent highly tentative approaches to the study of personality differences among cultural groups. Many are rather subjective and likely to reflect what the investigator expected to find. Another point to consider is that cultures are not homogenous. This is particularly true of modern nations, which represent a varied array of local "regional characters" in different sub-groups. As conditions change, moreover, "national character" may change. The descriptions cannot be expected to remain fixed, although certain features may persist.

GESTURE: AN EXAMPLE OF CULTURAL ASSIMILATION

It has often been maintained that racial groups manifest characteristic *bodily attitudes* and *movements*. The habitual postures, peculiar walk, and other traditional motor habits of various groups have been described at great length. Attention has also been called to the large group differences in the speed and tempo of movement. Special interest, however, has always been attached to the *gestural behavior* of

different peoples. The frequent emotional connotations of gestures, their peculiar relationship to language, and the easily observable differences in the traditional gesture patterns of various groups have made their study a particularly fascinating one. A voluminous literature has accumulated on this subject, most of the writings being either purely descriptive or speculative in nature. Artists, historians, philosophers, anthropologists, and many others have contributed their observations or theories to this topic (cf., e g., 11, 35). The layman, depending upon his mood and disposition, is amused, estranged, or repelled by the spectacle of a gestural pattern too unlike his own. In popular thought, gesture has been linked with underlying personality differences among racial groups. As a result, this phase of motor behavior has acquired a special significance in discussions of race differences.

A suggestive approach to the study of characteristic "racial" gestures is to be found in an investigation by Efron and Foley (11, 12). The groups employed were: (1) "traditional" Italians living in "Little Italy," one of the Italian districts in New York City; (2) "traditional" Jews living in New York's lower East Side; and (3) "assimilated" Italians and Jews, both living in similar "Americanized" environments. In view of the wide diversification in behavior patterns among different samplings of Italian and Jewish subjects, the authors further specify that the Jews included in this investigation were predominantly of Lithuanian or Polish extraction, and the Italians were from southern Italy, chiefly from the vicinity of Naples and from Sicily. The findings are thus restricted to these particular groups. Similarly, the results are to be qualified by the fact that only immigrant groups in America were employed.

The gestural behavior of these subjects was investigated by the following methods: (1) direct observation and description, (2) sketches made by an artist, and (3) motion pictures. All three methods were applied to gesticulation occurring in everyday life situations, the subjects being unaware of the fact that they were being observed. The motion picture material was subjected to two types of analysis. In the first place, the films were shown to naïve observers who were asked to judge various characteristics of the movements. The second method was more quantitative. The film, taken with a constant-speed moving picture camera, was projected frame by frame upon coördinate paper. The positions of motile parts, such as fingers, wrist, elbows, etc., were

marked in successive frame projections When these points were joined, a precise representation of the gestural behavior pattern was obtained Figure 95 illustrates this graphic technique in the case of a traditional Italian. It will be noted that there are four distinct lines of motion portrayed, the continuous lines representing the paths of movements of the right and left wrists, and the broken lines depicting the accompanying motions of the respective elbows The numbers indicate the direction of movement representing the position of the given part in each successive frame projection

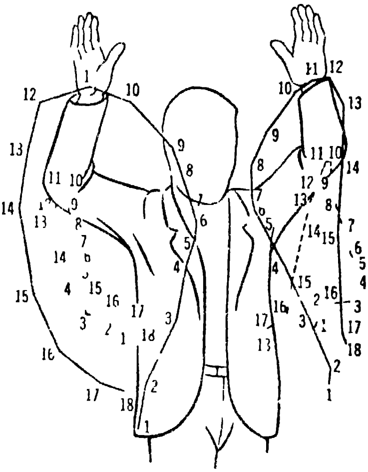


Fig. 95. Graphic Technique Employed in the Analysis of Gestural Behavior (From Efron and Foley, 12, p 14)

A study of the curves constructed by this technique as well as a consideration of the data collected by the other more qualitative methods, led to two principal conclusions First, clearly distinguishable and characteristic gestural patterns were exhibited by the traditional Italian and Jewish groups Some of the major differences between these patterns may be summarized as follows:

An analysis of the *parts of the body* involved in gesticulating revealed that the Italian tends to use preferably his arms whereas the Jew frequently employs his head as well as his arms hands and fingers, in a functionally differential way Head and finger gestures are rather typical of the Jewish expressive movements

The *form of the movement* also showed a marked contrast between the two groups In the Jew the movements are often sinuous and change direction frequently the Italian is more inclined to continue in the same direction until completion of the entire gesture segment

In regard to *laterality* (i e unilateral or bilateral) as well as *symmetry* of movement, pronounced differences were noted The Jewish gesture is predominantly asymmetrical, with frequent crossings and intertwinings Gesticulation is usually executed with one hand and arm, and if two are used they are employed in a sequential rather than a simultaneous fashion.

The Italian, on the other hand, frequently uses two arms simultaneously, and the movements are highly symmetrical in character.

The *radius* of the movement differed in the two groups, the Jew employing a relatively confined area, while the Italian sweep was found to be characteristically large, with movements involving the entire arm.

The two groups likewise varied in the *area in which gesticulation occurs*, the Jewish group seldom deviating from the medial plane of the body, whereas the Italian is more likely to perform his movements within the lateral areas.

Within each of these general areas, a difference was found in the *direction* of the gestural movements themselves, the Jewish movements being more frequently toward, and the Italian away from, the body of the gesturer.

Significant differences were likewise noted in *rhythm* or *tempo*, the Jewish movements being characteristically jerky, sporadic, and variable, while those of the Italians are more even and less variable.

The *frame of reference* of the gestures also differed. The Jewish gestures are more likely to be directed toward the body of the person addressed as a "*point de repère*," the speaker frequently touching the auditor, or literally "buttonholing" him. In contrast, the Italian gestures are typically oriented around the body of the speaker as a frame of reference

In addition to these spatio-temporal characteristics of the gestural movements themselves, certain major differences were observed in regard to the *meaningful* or *linguistic* function of such gestures. The Jewish gestures were characteristically of the discursive or logical type, being, as it were, a gestural portrayal not of the object of reference or thought, but of the process of ideation itself. This discursive or logical type is absent among the traditional Italians, whose gestures are frequently pictorial or pantomimic, the latter being a sort of re-enactment or imitation of the actions verbally described. Purely symbolic gestures are also common among the traditional Italian, and convey definite meaningful associations. These may be used to accompany verbal intercourse or may even function as the exclusive means of communication.

The second major point brought out by this investigation was that all the above characteristics of the traditional Italian and Jewish groups tended to disappear in the "assimilated" groups. In general, the more assimilated the individual, the less his gestural characteristics resembled those of traditional Jewish or Italian groups. The traditional differences between Jewish and Italian gestures were absent in the fully assimilated groups, and both resembled the particular "American" group with which they had become associated. On the whole,

gesticulation was much *less frequent* in such assimilated groups. The differences in gestural behavior between traditional groups and the lack of such differences between assimilated groups could not, furthermore, be explained on the basis of native or foreign birth. It was found, for example, that the American-born students at an orthodox Jewish school in New York City exhibited the gestural behavior of the traditional groups observed in the lower East Side, while the American-born Jewish subjects obtained at an exclusive Fifth Avenue club showed no such traditional gestures. In summary, a marked disparity was found between most of the gestural patterns characteristic of the traditional Jewish and Italian groups investigated, but no such contrasting gestural patterns were noted in assimilated groups of the same "racial" extraction. Thus cultural stimulation or habituation, rather than so-called racial descent, seems to be operative in the development of gesture.

CONCLUDING EVALUATION

In the two preceding chapters we have noted the many difficulties which beset the study of race differences in psychological traits. Race, defined as a biologically distinct group differentiated by common innate physical characteristics, is a difficult category to apply to contemporary man. In the attempt to arrive at a classification of human races, one proposed criterion after another has proved inadequate. An analysis of the major alleged physical differentia of race reveals wide variation within a single group, overlapping of groups, inconsistency with other criteria, and susceptibility to environmental influences. One or more of these criticisms can be leveled against each of the proposed criteria. Thus even the best possible classification of races is to be regarded as tentative and approximate. In fact, the very concept of race could be questioned on both theoretical and empirical grounds.

Race mixture, which has been going on for many generations, also adds to the complexity of the problem. The issue is further confused by the testing of immigrant groups which may not be representative samplings of their national populations. Moreover, immigrants are likely to be undergoing a period of intense readjustment and conflict arising from their contacts with the new culture, and this cannot fail to affect their behavior in many ways.

The problem of testing and comparing racial groups also presents

serious difficulties. Members of different races usually differ in many other respects as well. These differences often make direct comparison of behavior impossible. Thus language handicap has been shown to have a marked influence upon mental test performance. The subject's reaction to an examiner of a different race, the establishment of "rapport," the use of pantomime or of pictures which may not be equally familiar to all groups, all make the administration of tests a difficult task. The racial groups to be compared, furthermore, may not be equated in educational opportunities and facilities, socio-economic status, and the general cultural milieu in which they live. The special traditions, customs, and interests characteristic of each group may further "interfere" with test responses. Finally, it is impossible to establish a hierarchy of groups in terms of *absolute* intellectual superiority or inferiority. "Intelligence" tests measure certain abilities required for success in the particular culture in which they were developed. Cultures differ in the specific activities which they encourage, stimulate, and value. The "higher mental processes" of one culture may be the relatively useless "stunts" of another.

In so far as the members of different races live under varied cultural conditions, it is extremely difficult to compare them directly and impossible to determine the relative contribution of hereditary and environmental factors in producing any behavioral differences among them. In a few investigations, which have been reported in the present chapter, it was found possible to make cross-comparisons among racial and cultural groupings. In so far as these two categories, race and culture, cut across each other, it is possible to tease out the relative influence of biological and environmental factors. The results of such investigations are highly suggestive.

It would be premature, of course, to hazard any conclusive statements on so complex a problem, but the bulk of the evidence is definitely against the existence of behavioral differences among "races" in the biological sense. It is misleading to conclude that to date investigators have merely *failed to prove* race differences in behavior. The present state of our knowledge on this question is not a complete blank; nor is the evidence perfectly balanced, with half of the data favoring a racial hypothesis and half a cultural hypothesis. It is a fact that there are *group differences in behavior*, but not that such differences are racial or biological in origin. There is a considerable body of data, both in the racial studies and in other more general investigations

on the origins of individual differences in behavior, to show the influence of environmental factors in behavior development. But no study has conclusively demonstrated a necessary association between behavior characteristics and race as such.

To determine whether or not a behavior difference is truly racial logically implies three questions. First, is the behavior difference under consideration traceable to a structural difference? If so, is this structural characteristic gene-determined, i.e., not the result of dietary factors, birth injuries, or other environmental conditions? If both of the above questions are answered affirmatively, the final question is: Can a linkage be demonstrated between the genes determining this structural characteristic and the genes determining such racial characteristics as skin color, cephalic index, hair quality, and other commonly used criteria of racial classification? A negative answer to *any one* of these three questions precludes a racial interpretation of the observed behavioral difference.

CHAPTER

23

Socio-Economic Differences

ONE OF THE PRINCIPAL SHORTCOMINGS of most efforts to describe or understand "national character" is their tendency to gloss over important differences among cultural sub-groups within a nation. Moreover, the accounts are sometimes based, not upon the common features of the national culture, but upon an overgeneralized picture of the particular sub-group with which the investigator was most familiar. In America, such broad regions as the New England States, the South, or the Midwest will be readily recognized as differing in more than geography.

The distinction between "city" and "country" is likewise a familiar one. Even the casual observer is aware of significant differences between the urban and the rural dweller, not only in abilities, but also in interests, emotional responses, and general outlook. Actually this division is not a twofold one, but includes a series of groups, each differing from the others in distinct ways. From the large metropolis, through the moderately large city, the small town, the village with its one general store and post office, to the open country and the isolated mountain community, there are to be found many degrees and types of variation. The extremes of this series present definitely contrasting psychological pictures. Among the intermediate and more nearly adjacent members, there may not be a very pronounced intellectual variation, but in such cases well-known personality differences are often found. Thus the attitudes and emotional traits of the isolated mountain dweller and of the inhabitants of a small village may be fundamentally diverse. Similarly, between the resident of a large city and the member of a small town community there exist differences in outlook which have been repeatedly described and dramatized in literature.

Another kind of cultural grouping whose importance is receiving

increasing recognition is that represented by social classes. Sociological research in American communities has demonstrated not only the well-nigh universal prevalence of such social stratification, but also the profound effect which the individual's class membership may have upon his behavior development. The chief difference between a rigid "caste system" and the class systems found in a democracy such as that of the United States is the greater degree of "social mobility" possible in the latter.[1] Thus it is possible for the individual in a lower social class to rise to a higher status through his own efforts. It is this possibility which is at the root of many of the characteristic motivations and attitudes of the "middle class," with its emphasis upon hard work, self-improvement, and attainment.

An interesting practical application of the concepts of social stratification and class status is to be found in industry. The modern industrial psychologist recognizes that the plant personnel is structured into status groups or classes, in much the same way as any other community (cf. 76, 97). Not only occupational titles and wages, but distribution of working hours, characteristic wearing apparel and insignia, type of chair or desk, and almost any item or event in the working environment can become associated with these social distinctions and thus serve as a "prestige symbol." Any change which threatens to disrupt the individual's position in such a prestige scale may have a very demoralizing effect.

CLASS STRUCTURE AND PSYCHOLOGICAL DEVELOPMENT

The class differentiation of American society has been vividly demonstrated in a series of sociological studies conducted in certain typical American towns. These towns have become familiar by the pseudonyms given to them by the investigators: from the Middle West we have *Middletown* (58, 59), *Plainville, U.S.A.* (98), and *Prairie City* (38); from New England, *Yankee City* (94, 95);[2] and from the South, *Old City* (16). In all these studies, the research methods employed were similar to those developed by social anthropologists in their field

[1] It has been pointed out that the social position occupied by such ethnic minority groups as the American Negro is more nearly that of a caste than that of a class. Even within such a "caste," however, a further class stratification based primarily upon socio-economic factors is found (cf. 16, 19).

[2] The entire "Yankee City Series" comprises six volumes, four of which have appeared to date (94, 95, 96, 97); only the first two volumes are primarily concerned with class differentiation.

studies of preliterate cultures. The investigators lived for an extended period in the particular town, taking part in its social activities. Many local residents of varied socio-economic level were interviewed and their behavior in social situations was observed. By such techniques, information was obtained not only on the prevailing class concepts and criteria, but also on the social status of specific persons in relation

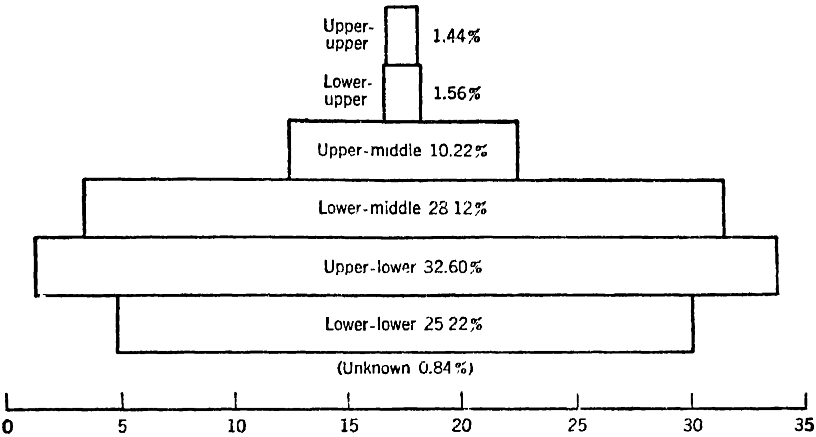


Fig. 96 Social Stratification in an American Community: "*Yankee City.*" (From Warner and Lunt, 94, p. 88.)

to other persons. The investigators wanted to discover "who associates with whom," and in what capacity. It was primarily on the basis of such information regarding social participation that the status classification of each individual was determined. Once this had been done, it was then possible to check such characteristics as income, property, education, church and club membership, and other factors which might differentiate the status categories. In general, the results of these surveys indicate a stratification into three major classes, each being further subdivided into two sub-classes, as shown in Figure 96. The per cent of persons falling into each of the six categories within a sampling of 16,785 persons investigated in *Yankee City* is also indicated in the figure. The relative proportion of persons in each status class did not differ substantially in the other towns studied.

These class distinctions were based largely on occupation and income level, although such factors as family background, education, beliefs and attitudes, and moral standards provided additional criteria.

The distinction between upper-upper and lower-upper in the New England and southern towns was made primarily in terms of family background, the upper-uppers representing the "old aristocracy," and the lower-uppers the "newly rich." In the midwestern communities, this distinction was not generally made, there being only one "upper" class comprising the wealthiest and most prominent families. The upper-middle class consisted principally of business and professional people, the "pillars of society," while the lower-middle class included small tradesmen, "white collar workers," and some skilled labor. The upper-lowers, consisting largely of semi-skilled and unskilled workers, were often described by middle-class persons as "poor but respectable" and "hardworking people." In contrast, the lower-lowers were characterized as shiftless and disorderly. The upper three classes together constituted "the big people" of the town, the lower three classes being regarded as "the little people."

An interesting by-product of these studies was provided by the comparison of class concepts held by individuals in different levels of the social hierarchy. Figure 97 illustrates these differences as found in the population of *Old City*. It will be noted that groups more remote from the informant are sometimes classed together, but finer distinctions are made in the vicinity of the individual's own class. The sharpness of the various divisions also differs with the group to which the informant belongs, as indicated by the number and position of the solid and broken lines in the diagram. The most conspicuous finding, however, is the frequency with which derogatory terms are used when describing other people's classes, and laudatory terms when describing one's own. The same class looks very different when viewed from the top, the middle, or the bottom of the ladder!

Of special interest to the differential psychologist are the effects which social class membership may have upon the individual's emotional and intellectual development. In his analysis of personality, Murphy (67) has maintained that the social classes show distinct "psychological cleavage" or discontinuity, and that these cleavages are reflected in personality structure. All surveys have corroborated the fact that these classes represent distinct cultural units. The type and extent of social contact between the various classes is definitely restricted. Moreover, the class stratification is reflected in large differences in home life, education, recreational outlets, reading habits, religious observance, and political activity.

UPPER-UPPER CLASS

UU — **"Old aristocracy"**

LU — "Aristocracy," but not "old"

UM — "Nice, respectable people"

LM — "Good people, but 'nobody'"

UL / LL — "Po' whites"

LOWER-UPPER CLASS

UU — "Old aristocracy"

LU — **"Aristocracy" but not "old"**

UM — "Nice, respectable people"

LM — "Good people, but 'nobody'"

UL / LL — "Po' whites"

UPPER-MIDDLE CLASS

UU — "Society" { "Old families"

LU — "Society" { "Society" but not "old families"

UM — **"People who should be upper class"**

LM — "People who don't have much money"

UL / LL — "No 'count lot"

LOWER-MIDDLE CLASS

UU / LU — "Old aristocracy" (older) | "Broken down aristocracy" (younger)

UM — "People who think they are somebody"

LM — **"We poor folks"**

UL — "People poorer than us"

LL — "No 'count lot"

UPPER-LOWER CLASS

UU / LU / UM — "Society" or the "folks with money"

LM — "People who are up because they have a little money"

UL — **"Poor but honest folk"**

LL — "Shiftless people"

LOWER-LOWER CLASS

UU / LU / UM — "Society" or the "folks with money"

LM — "Way high ups," but not "Society"

UL — "Snobs trying to push up"

LL — **"People just as good as anybody"**

Fig. 97. The Social Perspectives of the Social Classes: *"Old City."* (From Davis, Gardner, and Gardner, 16, p. 65)

Also relevant in this connection are the data collected by Kinsey, Pomeroy, and Martin (47) on male sexual behavior. On the basis of their intensive interviews with 6300 American men, the investigators were strongly impressed by the relationship between socio-economic level and sexual behavior. For example, the lower socio-economic classes report a higher incidence of pre-marital and extra-marital sexual relations than the higher socio-economic classes, but masturbation is more frequently reported in the higher socio-economic levels. Upper-class males also respond erotically to a wider range of stimuli than do lower-class males. The investigators themselves regard such socio-economic differences as one of the basic findings of their survey. They write:

> The data now available show that patterns of sexual behavior may be strikingly different for the different social levels that exist in the same city or town, and sometimes in immediately adjacent sections of a single community. The data show that divergencies in the sexual patterns of such social groups may be as great as those which anthropologists have found between the sexual patterns of different racial groups in remote parts of the world. There is no American pattern of sexual behavior, but scores of patterns, each of which is confined to a particular segment of our society (47, p. 329).

To be sure, these results may reflect no more than the degree of willingness or reluctance of American men in different socio-economic classes to *report* certain sexual activities. Even if this is the case, however, the data would indicate certain socio-economic differences in *attitudes* toward various forms of sexual behavior.[3]

A number of investigators have called attention to differences in *child-rearing practices* among social classes. Davis and Havighurst (17) studied this question by means of intensive interviews of upper-middle and upper-lower class families in Chicago. The interviews covered such matters as feeding schedules, toilet training, daytime naps, going out alone, hour at which child is required to be home at night, and age at which the child is expected to assume various responsibilities. Several statistically significant differences were found within both the white and Negro groups studied. The differences were such as

[3] The results of this study should also be qualified in the light of certain *selective factors* which tended to make the samplings unrepresentative of the general population. However, since the reported differences between socio-economic groups are so large, it is unlikely that the general conclusion cited above would have been substantially altered by the use of more highly representative samples.

to suggest that middle-class parents tend to be more rigorous in their child-training practices, frustrate the child more in feeding and cleanliness training, and expect children to take responsibility earlier. The authors are of the opinion that these class differences in child-rearing practices may affect subsequent personality development. Other investigators have shown that the *language development* of children is closely related to socio-economic status (cf. 63).

Davis (14, 15) has repeatedly discussed the many discrepancies in the type of training received by children of different socio-economic classes and the possible implications of these inequalities for intellectual and emotional development. Such differences range all the way from the eating habits and the type of clothing worn on different occasions to the choice of playmates and the individual's educational and vocational goals. Davis further maintains that the public schools are primarily adapted to the middle-class culture, since educational personnel is recruited principally from the middle class. This situation, according to Davis, makes the curriculum, type of incentives, and other aspects of the educational experience provided by the schools unsuited to lower-class children. He suggests that this may be an important reason for the frequent school maladjustment and educational backwardness of these children. The evidence does show that *school achievement* is positively correlated with socio-economic status (cf. 30, 31, 32).

Surveys by means of personality tests, questionnaires, and opinion-polling techniques have tended to substantiate the class differences which would be expected on the basis of cultural differentials. On *neurotic inventories*, school children from lower socio-economic levels have shown more evidence of maladjustment than those from middle and upper levels (6, 7, 30). Moreover, these class differences were found to be larger and more reliable than differences between native and foreign groups, or between urban and rural groups, tested in the same investigation. When groups of comparable socio-economic level were selected, the national and urban-rural differences tended to disappear.

In another investigation (60), children of professional fathers were found to be more dominant, extroverted, and emotionally stable, whereas children of skilled laborers had more worries. Some data are also available which suggest possible socio-economic differences in adolescent "prestige factors" and in the attitudes of adolescents

toward their age-mates (cf. 1). An intensive investigation of 16-year-olds in *Prairie City* (38) likewise indicated the role of social class in adolescent character development.

It should be noted that, when comparisons are made between *selected groups* from different social classes, the class differences may be obscured by the differential operation of selective factors (cf. Ch. 20). Thus in a comparison of urban and rural college women on a personality inventory, no significant difference in total adjustment score was found between the two groups (75). It may well be, however, that those rural girls who go to college are the very ones who most nearly resemble urban girls in their behavior, and whose cultural background has been most similar to that typical of an urban environment. It is interesting to observe that, despite such probable selective factors, *certain items* still showed large differences between the two groups. Moreover, the extreme scores, indicative of the best and poorest adjustment in the sample studied, were found in the urban group.

The same type of differential selection probably operated in an intensive study of personality and economic background conducted by Davidson (13). Several standardized personality tests were administered to 102 children between the ages of 9 and 12, whose IQ's ranged from 120 to 200. Socio-economic status, as determined by family income level, showed no significant relation to the large majority of personality indices employed in the study. Again it may be argued that children who score so high on intelligence tests represent a different selection from the upper than from the lower social levels. Such children may have been exposed to more nearly similar environments than would be true of the different social classes in their entireties. In this study, too, the differences were reduced but not eliminated by selective sampling. Certain characteristics did show a significant relationship to income level. Among such characteristics were reading preferences and habits, recreational preferences, possession of fears, and "liberalism" in social issues.

A direct and thorough approach to status differences in personality is represented by the investigations of Gough (31, 32) on high school students. Within a group of 223 high school seniors in a midwestern city, two extreme socio-economic samples were chosen on the basis of scores on the Sims Score Card for Socio-Economic Status.[4] An

[4] For a further discussion of this scale, cf. pp. 801 802.

item analysis, based on the responses of these two samples on the 550 items of the Minnesota Multiphasic Personality Inventory, revealed 34 items which yielded significant socio-economic differences (31). An examination of these items suggests that students of higher socio-economic level show stronger literary and artistic interests, have more social poise, security, and confidence in themselves and others, report fewer fears and anxieties, display more "emancipated" and 'frank' attitudes in moral, religious, and sexual matters, and are inclined to be more positive, dogmatic, and self-righteous in their opinions.

The 34 differentiating items were grouped into a "status scale," from which the personality status scores of another group of 263 students were computed. These status scores correlated .50 with "objective" status scores based on characteristics of home background. Moreover, the correlations of the personality status scores with each of a number of other variables closely paralleled the pattern of correlations of home status with the same variables. The variables with which each of these two types of status scores were correlated included each of the other scales of the Minnesota Multiphasic Personality Inventory, as well as other personality tests, intelligence and achievement tests, and academic grades (32). These correlations further suggested that students of higher social status show more satisfactory social adjustment, less insecurity, and less social introversion than do lower status students. The comparison of personality status and objective status scores suggests interesting possibilities for the prediction of social mobility in individual cases. Thus discrepancies between the personality status score and objective status score may be related to the individual's tendency to rise or drop in the social hierarchy. For example, an individual with low objective status score but high personality status score might be more likely to go to college than one with low status scores in both respects. If this hypothesis is verified, it might help to explain the relatively small personality test differences between socio-economic groups which are found when selected populations are compared, as in the previously discussed studies.

It is now well known that occupational groups exhibit characteristic differences in *interests* not only in strictly vocational matters, but in almost all areas of everyday life activity. These differences are, in fact, the foundation upon which such tests as the Strong Vocational

Personal communication from Dr. H. G. Gough. Cf. also H. G. Gough, "A New Dimension of Status: III. Discrepancies between the St Scale and Objective Status," *Amer. Sociol. Rev.*, 1949, 14, 275-281.

Interest Blank have been constructed. An even more relevant finding reported by Strong (89) pertains to the marked differences in interest pattern found between different occupational *levels*. Strong has devised a special scoring key for measuring the occupational level (O. L.) of the individual's interests. This was done by selecting those items which differentiated most clearly among such levels. The O. L. score is an index of how "aristocratic" the individual's interests are, or how far they diverge from the interests typical of unskilled laborers. Not only does this score differ with the position of the individual's occupation in the socio-economic hierarchy, but within any one occupation it tends to be higher for those men whose work is of a managerial character.

Some of the largest and most consistent class differences have been found in *attitude surveys*. Nation-wide polling studies (11, 55) as well as more intensive investigations in local areas, such as New Haven, Connecticut (64), and Akron, Ohio (43), agree in finding higher socio-economic level to be associated with more conservative attitudes, and lower socio-economic level with more radical attitudes. As one might expect, individuals who already occupy a more favored position in the social ladder tend to favor the preservation of the *status quo*. In general, too, middle-class persons are more concerned with advancement along vocational and other lines, while the lower classes emphasize security (64).

One of the most carefully controlled surveys on the attitudes of different social classes was conducted by Centers (11). A representative cross-section of the adult white male population, totaling 1100 persons, was interviewed during the summer of 1945. The interviews covered attitudes with respect to various major economic and social issues, as well as with respect to class identification. Included was a battery of six questions designed to test the respondent's conservative-radical orientation on socio-economic issues. On the basis of the replies to these six questions, individuals were classified into five categories in reference to expressed conservatism-radicalism. In Figure 98 will be found the relative frequency of these five response categories among individuals of different occupational levels. Separate results are given for urban and rural samplings. The occupational differences are large and clear-cut, the author concluding that such differences leave little doubt that people's politico-economic orientations are closely associated with their socio-economic statuses. Another interesting observation was that, within any single occupational category, those persons

who subjectively identified themselves with the "working class" expressed more radical attitudes than those who classified themselves in the "middle class."

In view of the large differences in the traditional activities, motivations, and attitudes of the various socio-economic classes, we would expect certain concomitant differences in *intellectual development*. In actual fact, nearly every investigation in which intelligence tests have

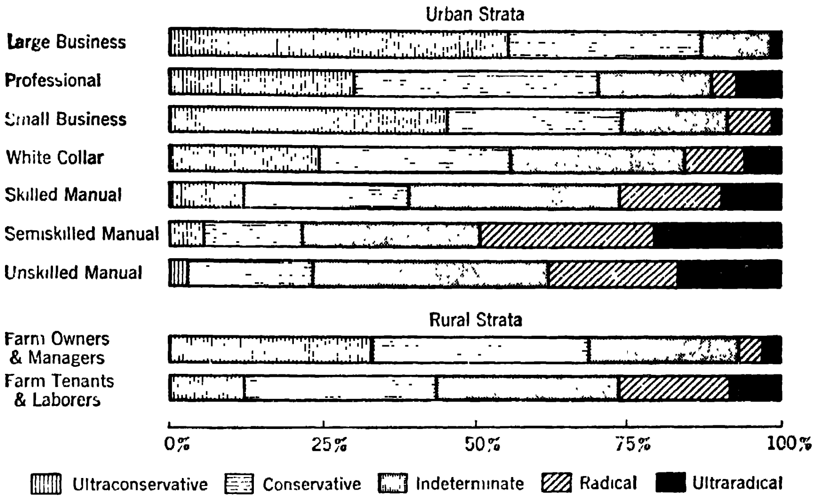


Fig. 98. Attitude Differences of Occupational Strata with Reference to Politico-Economic Conservatism-Radicalism. (From Centers, 11, p. 58.)

been administered to persons in different socio-economic levels has shown differences in the same direction. The trends are exceptionally consistent. In the sections which follow, we shall examine typical results obtained when intelligence test scores have been analyzed with reference to occupational categories, as well as with reference to other characteristics of social class. We shall also consider specifically some of the findings on relatively isolated groups, such as mountain dwellers, and on urban and rural populations as a whole.

OCCUPATIONAL LEVEL AND INTELLIGENCE

The first large-scale survey of the intelligence test performance of men engaged in different occupations was provided by an analysis of the Army Alpha scores obtained in World War I. On the basis of the

scores of about 18,000 men, the mean and range of Alpha scores were computed for 96 major occupations (24). The results fell into

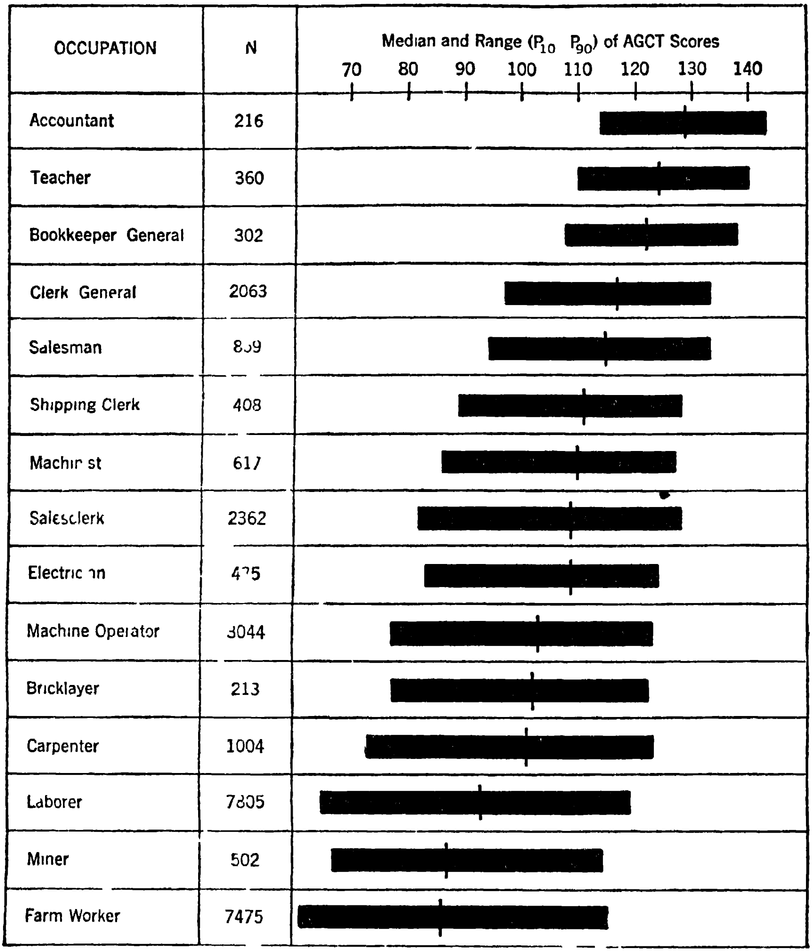


Fig. 99. AGCT Score in Relation to Civilian Occupation (Data from Stewart, 88, pp 5 13)

a distinct "*occupational hierarchy*," with the highest scores in the professional groups and the lowest among unskilled laborers Although overlapping was large and mean differences between adjacent groups

were slight, the differences between occupational groups in different portions of the hierarchy were large and statistically significant.

A corresponding analysis has been carried out with the AGCT scores obtained in World War II (88). Data on fifteen common occupations, selected from different parts of the hierarchy, are portrayed in Figure 99. In the complete analysis, similar data are provided for 227 occupations which were represented in sufficiently large numbers to yield reliable information. The results were based on the scores of 81,553 white enlisted men, derived from a random 2% sample of army personnel.[6] Because of selective factors operating in deferments, rejections, and discharges, such army samplings cannot be regarded as representative of adult civilian populations. Moreover, no officers were included in the occupational survey, thus further restricting the distribution at the upper levels. The representation of professional groups is especially limited by these factors. Data on doctors and engineers, for example, are virtually non-existent in such tabulations. Despite these limitations, it is clear that the intelligence test scores followed the general occupational hierarchy, paralleling the socio-economic status of the various groups.[7]

It may be of interest to note parenthetically that the intelligence test rank of those occupations which could be directly compared in the World War I and World War II figures correlated .84 (88). On the whole, the hierarchy remained substantially the same over the twenty-five-year period. None of the major occupational differences was reversed. Many of the small shifts which did occur, however, could not be attributed to chance fluctuations of sampling, but represented real differences in the populations tested during the two wars. A multiplicity of conditions, both in the army situation and in society at large, could account for such shifts in relative occupational intelligence. One possible factor may be found in shifts in the prestige value and in the skill requirements of certain jobs, as a result of technological advances. For our present purpose, however, it is the consistency of the major occupational differences, rather than the minor shifts, which is of primary concern.

[6] The 2% sample was a completely random sample of the entire army personnel In the occupational analysis, officers, enlisted women, and non-white enlisted men were not included, because of small number of cases or lack of AGCT scores As a result, the cases actually used represented approximately 1.6% of the total army population.

[7] A similar analysis conducted by Harrell and Harrell (75) on Army Air Forces personnel yielded results which are in substantial agreement with those obtained in the larger and more representative sample discussed above.

SOCIAL STATUS AND THE INTELLIGENCE OF CHILDREN

The correspondence between intelligence and occupational status is by no means limited to adults, but is also found among the children of men engaged in different types of work. Thus the relationship cannot be attributed primarily to differences in vocational experiences and in amount of formal education. More general conditions must be involved, which characterize not only the men, but also their families. The differences persist even when children in the same school grade are classified according to the occupations of their fathers. Such findings have been obtained in a large number of investigations, from the preschool to the college level.[8] In Table 59 will be found an occupational analysis of data collected during the standardization of the 1937

TABLE 59 *Mean Stanford-Binet IQ's of 2757 Children Classified According to Paternal Occupation*

(From McNemar, 66, p. 38)

	Father's Occupation	*Chronological Age of Child*			
		2-5½	6-9	10-14	15-18
I.	Professional	114.8	114.9	117.5	116.4
II.	Semi-Professional and Managerial	112.4	107.3	112.2	116.7
III.	Clerical, Skilled Trades, and Retail Business	108.8	104.9	107.4	109.6
IV.	Rural Owners	97.8	94.6	92.4	94.3
V.	Semi-Skilled, Minor Clerical, and Minor Business	104.3	104.6	103.4	106.7
VI.	Slightly Skilled	97.2	100.0	100.6	96.2
VII.	Day Labor, Urban and Rural	93.8	96.0	97.2	97.6

revision of the Stanford-Binet (66). These results, based on one of the largest and most representative samplings of American children ever tested, are typical of those found by other investigators with a variety of intelligence tests. In general, there seems to be a difference

[8] Many of these studies have been summarized by Neff (68) and by Loevinger (57).

of about 20 points between the mean IQ's of the children of professional people and those of the children of day laborers.

It should also be noted that these differences in IQ are just as conspicuous at the youngest age level (2–5½) as they are at the oldest (15–18). This finding is supported by other investigations within these age ranges (34, 71, 72). Moreover, studies on the intelligence of preschool children have revealed a similar relationship with paternal occupation. This is illustrated in Table 60, which gives the mean Kuhlmann-Binet IQ's of 380 children between the ages of 18 and 54 months, classified according to father's occupation (28). This group was retested within about six weeks, the results of both tests being shown in Table 60. The group differences were even larger on the retest than on the initial test.

TABLE 60 *Mean Kuhlmann-Binet IQ's of 380 Preschool Children Classified According to Paternal Occupation*

(From Goodenough, 28, p. 287)

	Father's Occupation	*Number of Cases*	*Mean Kuhlmann-Binet IQ*	
			First Test	Second Test
I.	Professional	56	116	125
II.	Semi-Professional and Managerial	29	112	120
III.	Clerical and Skilled Trades	129	108	113
IV.	Semi-Skilled and Minor Clerical	79	105	108
V.	Slightly Skilled	48	104	107
VI.	Unskilled	39	96	96

A number of investigators have employed scales for rating the socio-economic level of the home, thus permitting the computation of the *correlation* between each individual's intelligence test score and socio-economic level. Most of these scales involve visits to the home and interviews with parents, in order to obtain information for rating a number of home conditions.[9] One of the earliest of such scales was the Whittier Scale for Grading Home Conditions (101). Another, more restricted in scope but correlating highly with other scales, is the Chapin Living-Room Equipment Scale (12). The Sims Score Card for Socio-Economic Status (84) is a questionnaire covering cultural,

[9] For a survey of such scales, cf. Leahy (51) and Loevinger (57).

economic, educational, and occupational status of the family, to be filled out by the child himself. One of the most comprehensive and carefully standardized scales is the Minnesota Home Status Index, devised by Leahy (51). This scale yields a "home status profile" in terms of sigma-scores on each of six measures: children's facilities, economic status, cultural status, sociality, occupational status, and parental education. Each of the first four measures is based on from eleven to thirteen questions asked of one of the parents.

It should be noted that the correlation between socio-economic indices and intelligence is likely to be curvilinear, since the distribution of intelligence test scores is approximately normal, while the distribution of socio-economic level is quite skewed, with a piling of cases in the lower portion.[10] Consequently, computation of the usual Pearson coefficient of correlation between these two variables will underestimate the relationship between them. A few investigators have computed *eta*, or the correlation ratio, for this reason. Another solution is illustrated in the University of California Socio-Economic Index (4), in which family income is transmuted into a logarithmic scale, yielding a more nearly normal distribution; the transmuted income measure is then combined with amount of parental education, occupational level, and a composite rating of home, living room, and neighborhood.

Rather than treating socio-economic level as a unitary variable, some investigators have stressed the importance of determining what *specific features* of the environment are associated with intelligence test performance. In this connection, attempts have been made by Van Alstyne (93), Skodak (87), and others to design scales which cover the psychologically more significant aspects of the child's environment, such as parent-child contacts and the opportunities for various types of activity. A weakness of these scales is the role of subjective factors in the original choice of variables, as well as in the ratings themselves. However, they suggest interesting possibilities for further research.

It is apparent that the available indices of socio-economic level vary considerably in the aspects of environment which they measure. In view of the fact that the tests used to measure intelligence have also varied from one investigation to another, as has the choice of sub-

[10] Cf., e.g., the percentage of people in each of the social classes shown in Figure 96.

jects, it is not surprising to find a wide range of values given for the correlation between "socio-economic level" and "intelligence." Between the ages of 3 and 18 years, most of these correlations are in the vicinity of .40, although some are as low as .20 and some slightly over .50 (cf. 57, 68). The correlations show no consistent age trend within these age limits. Below age 3, the correlations drop (4, 41), and between birth and 18 months they are generally zero or slightly negative (4, 25). It should be recalled that psychological tests for infants are largely measures of simple sensori-motor development. As the child grows older, the tests become increasingly verbal and abstract in content. Since different functions are tested, there is thus no real inconsistency between the correlations obtained on infants and those on older children. On the whole, studies employing the correlation technique have corroborated the cruder results obtained by the comparison of occupational groups.

Approaching the same problem from a slightly different angle, Havighurst and his co-workers (36, 37, 42) made use of the "*social status method*" discussed in the first section of the present chapter. A midwestern community [11] with a population of about 10,000 was chosen, and the families placed on a scale of five social classes by the previously described methods. Psychological tests were administered to nearly complete samplings of children at three age levels, 10, 13, and 16. The 10- and 16-year-olds were given well-known intelligence tests, including verbal and performance types, as well as specialized tests of reading and of spatial and mechanical aptitudes. The 13-year-old group was tested with the Thurstone Tests of Primary Mental Abilities. The mean scores of the different status groups on each of these tests are summarized in Tables 61, 62, and 63, shown on the two following pages.

These three surveys are somewhat limited by the small number of cases available at each status level. The representation of the two highest social classes (A and B) was especially inadequate: in the 10- and 13-year-old samples there were too few A or B children to warrant the inclusion of these categories in the statistical analyses; and in the 16-year-old sample there were only 9 children in both classes combined. Despite the small number of cases, nearly all tests in all three samples show a *tendency for scores to rise with social status.*

[11] Called *Midwest* in these articles but elsewhere designated by the pseudonyms of *Prairie City* (38) and *Elmtown* (40).

TABLE 61 *Mean Test Scores Obtained by 10-Year-Old Children in Different Social Status Groups*

(Adapted from Havighurst and Janke, 37, p. 363)

Social Status	*N*	*Stanford-Binet IQ*	*Cornell-Coxe IQ*	[illegible] *Non IQ*	*Iowa* [illegible] *Quotient*	*Minn. Paper Form Board Score*	*Porteus Maze MA*	*Minn. Mech. Assem. T-Score (Boys)*	*Chicago Mech. Assem. T-Score (Girls)*
C. Lower-middle	26	114	116	107	99	22.5	12.7	52.5	56.0
D. Upper-lower	68	110	110	102	99	21.3	12.8	49.2	49.5
E. Lower-lower	16	91	96	91	88	15.7	10.4	46.9	41.3

TABLE 62 *Mean Scores on Thurstone Tests of Primary Mental Abilities Obtained by 13-Year-Old* Children in Different Social Status Groups*

(Adapted from Havighurst and Breese, 36, p. 244)

Social Status	*N*	Thurstone Scores					
		Number (*N*)	Verbal Comprehension (*V*)	Space (*S*)	Word-Fluency (*W*)	Reasoning (*R*)	Memory (*M*)
C. Lower-middle	21	52.7	55.6	52.5	53.4	53.8	52.5
D. Upper-lower	58	51.6	50.0	50.3	50.4	49.9	50.0
E. Lower-lower	11	36.4	39.4	43.0	40.1	43.4	44.1

*This group was studied later than the 10-year-olds and 16-year-olds.

TABLE 63 *Mean Test Scores Obtained by 16-Year-Old Children in Different Social Status Groups*

(Adapted from Janke and Havighurst, 42, pp. 503-504)

Social Status	*N*	*Stanford-Binet IQ*	*Wechsler-Bellevue Perf. IQ*	*Iowa Silent Reading T-Score*	*Minn. Paper Form Board Score*	*Minn. Mech. Assem. T-Score (Boys)*	*Chicago Mech. Assem. T-Score (Girls)*
A. Upper / B. Upper-middle	9	128	118	58.0	44	46.8	62.1
C. Lower-middle	44	112	109	51.0	40	51.6	52.0
D. Upper-lower	49	104	102	48.9	31	48.8	48.5
E. Lower-lower	13	98	103	45.5	31	53.0	45.3

To be sure, some of the differences between adjacent status groups are insignificant and a few are even reversed; but when extreme groups are considered, most of the differences are statistically significant.

The most clear-cut exception to the above trend is found in the Minnesota Mechanical Assembly Test given to the 16-year-old boys (Table 63). The mean scores on this test showed no consistent trend in relation to social status, and the highest mean was obtained by the lowest social group. A possible reason for these discrepant results is that the lower-status boys may have had more experience in dealing with mechanical objects and may thus have been more familiar with the tasks involved in the test. Direct comparisons among different tests are complicated by the fact that their units may not represent comparable steps, even when expressed as IQ's. It is nevertheless possible to detect a *tendency for status differences to be more conspicuous in the more highly verbal type of tests*. For example, among the 16-year-olds, the mean difference between extreme status groups in Wechsler-Bellevue Performance IQ was only 15 points, as compared to 30 points in the Stanford-Binet. The critical ratios were 2.4 and 4.1 for the former and latter differences, respectively. Moreover, the position of groups D and E is reversed in the Performance IQ, while all groups vary in a consistent direction on the Stanford-Binet. Similarly, among the 10-year-olds, the mean IQ differences between extreme status groups are 23, 20, and 16 IQ points on the Stanford-Binet, Cornell-Coxe, and Goodenough tests, respectively. Although the status differences on the different tests are more nearly uniform among the 10-year-olds than among the 16-year-olds, even in the younger group the Stanford-Binet tends to show larger critical ratios than do the performance scales.

It might be noted parenthetically that both age and sex comparisons in this investigation provide interesting corroboration of some of the points discussed in the chapter on trait organization (Ch. 1). Thus among 10-year-olds, all tests correlated fairly highly and uniformly with each other. Among the 16-year-old *boys*, mechanical aptitude appears to have become differentiated as a special aptitude, but this is not true for the 16-year-old *girls*. The status differences also fit in with these developmental changes. Thus among the 16-year-old boys, mechanical aptitude did not show the same status differences as did the verbal tests. Among the girls and among the younger children of both sexes, the status differences in mechanical aptitude were more

nearly similar to those in verbal tasks. This would be expected, since mechanical aptitude was not so clearly differentiated from verbal aptitude within these two groups.

In the Thurstone tests administered to the 13-year-olds, it is again apparent that status differences vary with the function tested. The mean scores of the status groups are shown in Table 62. None of the differences between groups C and D was significant at the .01 level of confidence, although the largest critical ratio (2.0) was found on the Verbal Comprehension test. When level 1 is compared with either C or D, the critical ratios for the Number, Verbal Comprehension, and Word Fluency tests range from 2.8 to 4.9, while those for the Spatial, Reasoning, and Memory tests range from 1.5 to 2.5. The same discrepancy is indicated by the correlations between score on each of the six tests and "index of status characteristics." The latter is a composite index based upon occupation, source of income, house-type, and community-area in which home is located. The investigators report that in *Midwest* this index agreed closely with the social position of individuals as found by the more elaborate social status method. The correlations of the status index with scores on each of the Thurstone tests were as follows:

Number (*N*)	.32 ± .10
Verbal Comprehension (*V*)	.42 ± .09
Space (*S*)	.25 ± .10
Word Fluency (*W*)	.30 ± .10
Reasoning (*R*)	.23 ± .10
Memory (*M*)	.21 ± .10

It will be seen that the correlations with the Number, Verbal Comprehension, and Word Fluency tests are higher and more nearly significant than those with the remaining three tests. The investigators suggest that the correlations with social status are higher in those abilities which might be favored by a superior social environment.

Some investigators have demonstrated a correspondence between the socio-economic ratings of a whole *community* and the mean IQ of its children. In a study of over 300 neighborhoods in New York City, each with a population of about 23,000, Maller (61) found a correlation of .50 between economic status of the neighborhood and mean IQ of its school children. Economic status was determined from federal census data regarding value of home rentals in the neighborhood; children's IQ's were based upon a battery of group tests, including

the National Intelligence Test and the Pintner Survey Test, given to over 100,000 fifth grade public school children. Working with even broader units, Thorndike and Woodyard (92) report very high correlations between the mean National Intelligence Test scores of sixth grade pupils and various social indices obtained for 30 cities. For example, the intelligence test scores correlated .78 with the index of per capita income for each city, and they correlated .86 with a composite index of the general "goodness" of community life, based on a variety of criteria.

Mention may also be made of studies conducted in *other countries,* all of which demonstrate the same correspondence between intellectual and socio-economic variables. Whether children are classified into a few categories on the basis of paternal occupation, or whether more precise socio-economic indices are correlated with intelligence test scores, the results closely corroborate those obtained on American children. Comparable data have been obtained on large groups of subjects from early infancy to high school age in such countries as England (10, 21, 65), Scotland (23), Poland (69), Rumania (70), the Soviet Union (20, 85), and Hawaii (56).

In all these comparisons of intellectual and socio-economic variables, we must not lose sight of the wide range of individual differences *within each level,* nor of the related fact of *overlapping between levels.* The fact that correlations between individual test scores and individual ratings for socio-economic factors fall far short of 1.00 is just another indication of this overlap. It is also well to remember in this connection that the total number of persons in the lower socio-economic or social-status classes is larger than that in the upper levels. The result is that if we begin with intellectual rather than with socio-economic categories, we may find that a larger percentage of intellectually superior persons come from the lower than from the upper social classes. For example, in a survey of more than 100,000 high school seniors in Wisconsin, the investigators report the occupations of the fathers of those students who fell above the group median in intelligence test score (8). Within this sub-group, only 7.9% had fathers in the professions, while 17.4% had fathers in skilled labor. This was true despite the fact that the median percentile score of all subjects with professional fathers was 68.5 and that of all subjects with fathers in skilled labor was 51.1. These findings are by no means peculiar to this study. Similar results would be obtained in nearly every study, if the data

were expressed in the same form. This simply means that the lower socio-economic classes may contribute more individuals of fairly high intelligence than the upper classes, although relative to the total number of persons in each socio-economic class, the contribution of the higher socio-economic classes is greater.[12]

A word may also be added regarding possible *interpretations* of the relationship between socio-economic factors and intelligence. The association *per se* does not, of course, provide any clue regarding causation. On the one hand, it can be argued that the intellectual differences found today among social groups testify to a gradual hereditary differentiation which has been going on through selection. Thus the more intelligent individuals would gradually work their way up to the more demanding but more highly coveted positions, each person tending eventually to "find his level." Since intellectually superior parents tend to have intellectually superior offspring, the children in the higher social strata would be more intelligent, on the whole, than those from the lower social levels. A second hypothesis would explain the intellectual development of the child in terms of the cultural level in which he is reared. Thus the child who grows up in the home of a construction laborer does not have the same opportunities for intellectual development—and consequently will not reach the same ability level—as a child of equal initial capacity brought up in the home of an eminent scientist or author. A third possible hypothesis is that the relationship between socio-economic and intellectual variables is *indirect* rather than direct. Thus both sets of variables may be related through some other factor, such as personality characteristics, national origin, or family size (57).

We cannot choose among these hypotheses without probing further into the particular circumstances in each case. Some investigators have been impressed with the finding that class differences in intelligence test performance appear so early in life and are practically as large among 3-year-olds as among 18-year-olds. This has frequently been regarded as evidence for a hereditary interpretation of class differences, on the grounds that environmentally produced differences should increase with age, as environmental factors have more time to operate. It is impossible, however, to make a universal generaliza-

[12] Beyond a certain point, the larger N in the lower classes would no longer be sufficient to counteract the shrinking proportional contribution. This was true, for example, in Terman's group of gifted children, in which 31.4% had fathers in the professional class, as compared to 11.8% in the skilled labor class (cf. Ch. 17).

tion regarding the relation of age to environmental influences. The trend may well be opposite in different situations. If the environmental differences continue to an equal degree or increase with age, then we should expect their differentiating effects on behavior also to increase. But if any equalizing influences are introduced into the environment at certain ages, as at the time of school entrance, these might counteract the divergence of behavior development otherwise expected.

A comparison of class differences in various types of ability may throw some light upon the origin of such differences, especially when the results are examined in the light of cultural dissimilarities among the groups. In the sections which follow, additional data bearing upon these questions will be presented from investigations conducted on relatively isolated groups, as well as on urban and rural populations.

THE INTELLECTUAL DEVELOPMENT OF ISOLATED GROUPS

Certain groups have been of special interest to psychologists because of their relative isolation from outside social contacts. One of the most widely quoted studies on such isolated groups is that conducted by Gordon (29) on canal-boat and gypsy children in England. Gordon's report, made in the course of his official duties as Inspector of Schools, was based on the Stanford-Binet IQ's and educational test scores of various groups of children whose schooling was deficient. The *canal-boat children* were enrolled in special schools maintained for them, which they were able to attend only while the canal boats were tied up for loading or discharging. It was estimated that the average school attendance of these canal-boat children was only 5% of that in ordinary elementary schools. The majority were able to attend school only about once a month for one or two consecutive half-days. Their home surroundings, although satisfactory in respect to conditions of health and cleanliness, were intellectually of a very low order. Many of the adults were themselves illiterate, and each family led a relatively isolated existence, with a minimum of social intercourse.

The average IQ of the entire group of 76 canal-boat children was 69.6. Taken at face value, this would suggest at best a borderline group, with a few distinctly feebleminded individuals. Further analysis

of the data, however, brought out the fact that IQ *declined sharply with age* within the group, the 4- to 6-year-olds obtaining an average IQ of 90, while the oldest group (12 to 22 years) averaged only 60. The correlation between IQ and age was —.755. Even when children in the same family were compared, a consistent drop in IQ from the youngest to the eldest sibling was noted. Moreover, the mental ages of children within a single family tended to be very similar, even though their chronological ages differed. Such a mental age might well represent the limit of intellectual development which was made possible by the available educational opportunities and the type of home environment furnished within the given family.

The results on the *gypsy children* were similar but less extreme than those on the canal-boat group. The mean IQ of the 82 gypsy children was 74.5, and the correlation between age and IQ was —.430. Thus both the total inferiority and the age decrement in intelligence were less pronounced in this group than in the canal-boat group. Corresponding to these findings is the fact that the school attendance of the gypsy children averaged considerably higher than that of the canal-boat children, being 34.9% of the total number of possible school days. The gypsy families led a nomadic existence, the children attending school only during the few winter months when they had a fixed abode. Although their living conditions were crude and primitive, these gypsy children had more social contacts outside of their immediate family, and were thus less isolated than the canal-boat children. It is also noteworthy that within the gypsy group, IQ showed a significant positive correlation of .368 with amount of school attendance for each child. It is possible, of course, that the relationship between amount of schooling and intelligence, both within and between these two groups, may have resulted in part from the greater willingness of the brighter children to attend school regularly. At least two different factors, however, seriously weaken this hypothesis. In the first place, it was physically impossible for the children to attend school while the canal boats or gypsy caravans were in motion. Secondly, the gypsy children frequently had to be forced by local authorities to attend school during their brief winter periods of stable residence.

Of special interest is the age decrement reported by Gordon for both canal-boat and gypsy children, but not found in surveys of more

privileged groups. One possible explanation for such a decrement is that the intellectual needs of the younger child can be satisfied almost as well in the restricted environment of the canal boat or gypsy camp as in a prosperous urban home. As the child grows older, however, the differential effects of poorer home environment and of deficient schooling become increasingly apparent. Another factor which undoubtedly enters into the obtained results is the well-known difference in the functions measured by intelligence tests at the lower and upper age levels. The increasing emphasis upon verbal and other abstract functions at the older ages might well present a progressively greater handicap to children whose environments do not encourage the development of these abilities. To be sure, the data do not in themselves preclude an interpretation in terms of some hereditary structural deficiency which might make these particular groups inferior in verbal and abstract functions. It might be argued that such a deficiency would not be apparent at the younger age levels, since these functions cannot be adequately tested among young children.

On the basis of the data presented by Gordon, it is not possible to choose conclusively between these two hypotheses. It should be noted, however, that no hereditary structural basis for verbal aptitude or other functions measured by intelligence tests has yet been discovered, nor does its discovery appear likely or plausible in the light of what we do know regarding the mechanism of heredity. On the other hand, there is a multitude of known factors in the environments of these children to account for their deficiencies. In fact, it is difficult to see how any child, whatever his heredity, could obtain a normal or superior Stanford-Binet IQ if reared in the environments of these canal-boat or gypsy children.

Studies on *mountain children* have closely corroborated Gordon's findings. An unusually good opportunity for the study of isolated communities is offered by the highlanders of our southern mountains. Owing to poor roads and general inaccessibility, many of these groups live in complete isolation during the larger part of the year. In certain districts, the cultural level is extremely low, little more than the bare necessities of life being available. Families are frequently found living in the original crude huts built by their ancestors several generations ago. Racially these groups are relatively homogeneous, being predominantly of British descent. They are highly inbred, and in certain communities only two or three different surnames

are to be found. The peculiar customs and manners [13] of the southern mountaineer have long stirred the imagination of author and playwright. As a result these highland people have achieved a certain amount of glamour in the mind of the public, a kind of notoriety which overshadows the squalor of their lives. To the psychologist, these groups offer a challenging opportunity to unravel the forces of heredity and environment.

Intelligence test surveys of children living in such isolated mountain communities have been conducted by Hirsch (39) and Asher (2) in Kentucky, Sherman and Key (82) in the Blue Ridge Mountains, Edwards and Jones (22) in Georgia, and Wheeler (99, 100) in Tennessee. The results of all these studies are quite uniform. Average IQ is clearly below the national norms; the inferiority is more marked on verbal tests, such as the Stanford-Binet and National Intelligence Test, and less marked on non-language and performance scales; and the same type of age decrement reported by Gordon is found among these mountain children.

Typical results from the study conducted by Sherman and Key (82) are shown in Table 64. The subjects included 102 mountain children living in four "hollows" in the Blue Ridge Mountains, approximately 100 miles from Washington, D. C., as well as 81 children living at Briarsville, a small village situated at the base of the Blue Ridge. These subjects represented over one-half of all children living in the five centers. Each of the five communities differed in length of school term, quality of schooling, and general level of material culture. Racially, however, the subjects were quite homogeneous, all being descended from a common ancestral stock. It was thus possible to make intercomparisons among the groups, in addition to an evaluation of scores in terms of urban norms.

It will be noted from Table 64 that both village and mountain children fall below the "normal" IQ of 100 on nearly all tests. The inferiority is, however, less pronounced among the village children, who had better schooling facilities. Both groups show a fairly consistent age decrement, which is also less marked in the village group. In the case of the mountain children, the mean IQ's tended to be lower on the verbal than on the non-verbal and performance tests.

[13] For descriptive material regarding these people and their surroundings, the reader may examine the accounts by Campbell (9), Kephart (46), and Raine (74), as well as the more recent report by Lewis (53).

TABLE 64 *Mean IQ of Mountain and Village Children in Relation to Age*

(From Sherman and Key, 82, p. 287)

Age	*Pintner-Cunningham Test*		*National Intelligence Test*		*Goodenough Draw-a-Man Test*		*Pintner-Paterson Performance Tests*	
	Mt.	Vill.	Mt.	Vill.	Mt.	Vill.	Mt.	Vill.
6– 8	84	94			80	93	89	
8–10	70	91		117	66	82	76	93
10–12	53	76	66	101	71	69	70	87
12–14			67	91	69	73	83	
14–16			52	87	49	70	73	

Of special interest is the study conducted by Wheeler (99, 100) on East Tennessee mountain children. Group intelligence tests were administered in 1940 to over 3000 children in forty mountain schools. The results were compared with those obtained on children in the same areas and largely from the same families, who had been similarly tested in 1930. During the intervening ten-year period, the economic, social, and educational status of these sections is reported to have improved considerably. Paralleling such environmental improvements, a rise in IQ from the first to the second sampling was noted at all ages and all grades. The median IQ's were 82 and 93 in the 1930 and 1940 samplings, respectively. The usual age decrement was found in both samplings. In the 1930 group, the median IQ dropped from 94.7 at age 6 to 73.5 at age 16; in the 1940 sample, it dropped from 102.6 at age 6 to 81.3 at age 15.

It might be added that such an age decrement is not limited to the relatively unusual groups of children discussed in the present section, but has also been reported for other underprivileged groups. This decrement is especially apparent where educational facilities are deficient. Among the groups for which such a drop in IQ with age has been found may be mentioned: southern mill-town children of low socio-economic level (45); children reared in a "high delinquency area" in a large city (54); and children admitted to an orphanage after varying periods of residence in their own, very inferior homes (86). Such age decrements have also been noted in many investigations on rural children, to be discussed in the following section.

INTELLIGENCE TEST SURVEYS OF RURAL CHILDREN

The distinction between urban and rural populations is partly one of occupation, but it also involves other important aspects of the physical and social environment. Most of the differences are such as to handicap the rural child in academic progress and in the type of abilities sampled by most intelligence tests. Thus educational opportunities are notoriously poor in many rural districts of our country, in sharp contrast to the excellent facilities available in most towns and cities. The length of the school term is often shortened in rural communities because of the impassable condition of the roads at certain times of the year, or because the children are needed to help with farm duties in busy seasons, or for other reasons of a local nature. In some cases the school term lasts only six months. Similarly, the difference in type and amount of instruction received in the "consolidated" and the "one-room" school is a very real one. In the latter type of school, in which pupils of all ages and grades are taught by a single teacher and in a single classroom, progress must necessarily be very halting. Differences in the provision of books and other supplies, as well as in teacher training, are too obvious to mention.

The general cultural milieu of different localities likewise presents striking contrasts. Libraries, museums, and other facilities for the intellectual or artistic stimulation of the community are far more accessible and better developed in urban than in rural districts. The recreational activities of rural children are quite different from those of urban children, as shown, for example, in the extensive survey of play activities conducted by Lehman and Witty (52). These investigators concluded that the differences are "directly traceable to environmental opportunities," and that such differences may in turn influence the direction of the child's intellectual development. The extent and variety of social contacts also differentiate city and country groups. Between the cosmopolitan associations of the large metropolis, with its kaleidoscopic array of diverse customs, manners, and peoples, and the relatively homogeneous and sparse contacts of the rural village or open country, there exist tremendous differences in social stimulation.

The fact that country children as a group score distinctly below the city norms on current intelligence tests has been repeatedly demonstrated. Numerous investigations, some employing several thousand

children and covering practically complete school populations, have consistently revealed the inferior performance of rural children in all parts of the United States.[14] Typical results are given in Table 65, based upon McNemar's analysis of the standardization sample of the 1937 Stanford-Binet (66). The investigators made a very serious effort to obtain urban and rural samplings which were representative of the country at large, but they express a certain amount of skepticism regarding the adequacy of the rural sampling. They point out, however, that the selective factors were such as to reduce the urban-rural differences in their data. Thus the differences between city and country children would probably have been still larger if a more representative group of rural children had been surveyed.

TABLE 65 *Mean Stanford-Binet IQ's of Urban, Suburban, and Rural Children*

(From McNemar, 66, p. 37)

Locality	Age Range in Years					
	2–5½		6–14		15–18	
	N	Mean	N	Mean	N	Mean
Urban	354	106.3	864	105.8	204	107.9
Suburban	158	105.0	537	104.5	112	106.9
Rural	144	100.6	422	95.4	103	95.7

Separate means for urban and suburban groups are included in Table 65, but as would be expected, these groups did not differ appreciably. Suburban communities are within commuting distance of large cities, and they share most of the benefits of urban centers. The rural children, on the other hand, average about 10 IQ points lower than the urban during school age (6–18), and about 5 points lower during the preschool period (2–5½). It is noteworthy, too, that the investigators report a slight tendency for rural IQ's to drop at the beginning of the school period, no such tendency having been found among urban children (91).

That age is an important factor in the amount of urban-rural difference was also demonstrated in the thorough and comprehensive investigation conducted by Baldwin, Fillmore, and Hadley (3) on Iowa farm children. Children in four rural communities were com-

[14] For a good summary of studies conducted prior to 1930, cf. Shimberg (83).

pared with Iowa City children, as well as with the test norms. The number of rural subjects tested at each age level, together with the tests employed, are shown below:

123 infants (4 to 40 months)	Iowa Baby Tests
163 preschool children (3 to 6 years) . . .	Detroit Kindergarten Test
871 school children	Stanford Binet
	Otis Intelligence Test
	Pintner-Paterson (5 selected tests)

The results showed that, among the rural *infants*, there was no noticeable inferiority on the baby tests. Nor can this lack of differentiation be attributed to a deficiency in the discriminative power of the tests, since wide individual differences were obtained. In the *preschool* group, a rural inferiority appears in the 5- and 6-year-old groups, no significant differences having been found at the younger ages. The rural *school children*, however, showed a definite intellectual retardation which became increasingly large as they progressed through school.

This deficiency, futhermore, was more pronounced in the one-room than in the consolidated schools. In the latter type of school, the intellectually retarded children were found chiefly at the upper ages; whereas in the one-room schools, the median mental age was always lower than the median chronological age. The median mental age deficit in the one-room schools ranged from 1 to 6 months up to the age of 9; between the ages of 9 and 12 it increased from 7 to 14 months; and at ages 13 and 14 it amounted to 16 and 39 months, respectively. In the consolidated schools, on the other hand, the median mental age exceeded the median chronological age up to the age of 13, the excess for each age ranging from 1 to 8 months. From 13 to 18 years, the median mental age was lower than the median chronological age, the deficit in successive years being 2, 6, 19, 10, 11, and 10 months, respectively. The drop in amount of retardation beyond age 14 may be due to the more select nature of the older groups in rural schools.

An analysis of the rural children's performance on different tests or parts of tests revealed their handicap on verbal materials. In regard to the performance tests, it is interesting to note that the farm children excelled on the Mare-and-Foal Test, a picture completion test portraying a farm scene. On all other performance tests which involved *speed*, the rural subjects were deficient, their movements tend-

ing to be slow and deliberate. The usual instructions to work rapidly did not in themselves seem to provide sufficient incentive for these children. The rate of movement could, however, be increased if other appeals were added. The investigators suggested that "the children's apparent lack of comprehension of the meaning of hurry is to be expected as a consequence of some of the influences that surround them" (3, p. 254).

In summary, these studies, together with many other similar surveys, show rural children to be consistently inferior on both intelligence tests and educational achievement tests. This inferiority tends to be greater on verbal than on non-verbal tests, and greater on tests which emphasize speed. With increasing age, rural scores tend to decline in relation to urban norms. Rural inferiority is also more marked on group than on individual tests. It has been suggested that the country child's performance on a group scale may be handicapped by his greater shyness with strangers (71). This difficulty would be partly overcome by the examiner's efforts to establish rapport in the administration of an individual scale.

Average scores also tend to be lower in those districts with poorer schooling facilities. This is most clearly apparent in the comparison of one-room with consolidated schools. A somewhat different analysis of rural areas was employed by Pressey and Thomas, in their study of Indiana farm children (73). When school children were classified into those living in "poor" farming districts, where the land was hilly and the soil inferior, and those living on "good" farm land, the two groups were found to differ significantly in intelligence test performance. Among the children in the good farming districts, 36% reached or exceeded the median of city children, as contrasted to only 20% in the poor farming districts. When a younger group (ages 6 to 8 years) was tested in the poorer farming area, 22% of the children reached or exceeded the urban median (71). In explanation of these findings, the authors suggest that a constant selective process goes on in farming areas, the inferior, less intelligent families being pushed back into the hill country where the soil is poorest. It should also be noted, however, that in the poorer farming area studied by Pressey and Thomas, educational facilities were notoriously deficient and socio-economic level of the homes was lower than in the good farming area.

There is a considerable body of data from several European countries demonstrating the existence of urban-rural differences in intelligence test performance. Thus Klineberg (48), in the investigation discussed in the preceding chapter, found that the urban-rural differences in performance on the Pintner-Paterson tests were much larger than either racial or national differences. The relevant data are summarized in Table 66. The mean difference between the entire

TABLE 66 *Performance Test Scores of Urban and Rural Groups in Europe*

(From Klineberg, 48, p. 27)

Group	*Number of Cases*	*Mean*	*Median*	*Range*
Paris	100	219.0	218.9	100-302
Hamburg	100	216.4	218.3	105-322
Rome	100	211.8	213.6	109-313
Total city	300	215.7	216.9	100-322
Total country	700	187.1	187.0	63-314

urban sample of 300 children and the entire rural sample of 700 is over eight times as large as its standard error, and is hence clearly significant. In terms of overlapping, only 30.12% of the rural children reached or exceeded the median of the urban children. It is also interesting to note that the three city groups, tested in Paris, Hamburg, and Rome, differed little *among themselves*. None of the differences between these three city means is statistically significant. The rural groups, it will be recalled, revealed larger and fairly significant national differences. It would seem that the equalizing effect of life in a large cosmopolitan city tends to obliterate many of the differences arising from the specific national culture. Equally striking urban-rural differences were found by Rosca (77) in Rumania. On a series of locally constructed intelligence tests given to 2032 children, the mean IQ of the urban children was 107, that of the rural groups, 86.

A number of investigations conducted in Great Britain have shown much less urban-rural differentiation in intelligence test performance than has been found in America or in other European countries.

Especially is this true of the more remote rural districts of Great Britain, which frequently show no inferiority to the urban areas (5, 90). A particularly clear demonstration of this point is provided by the results of one of the Scottish Surveys, in which the Stanford-Binet was given to a complete sampling of all children born in Scotland on June 1, 1921 (cf. 78). The relative performance of urban and rural groups in this survey was closely corroborated by a more extensive survey with group tests. The Stanford-Binet results are shown in Table 67. Mean IQ's are reported for the four cities, the

TABLE 67 *Mean IQ's of Children in Urban and Rural Areas of Scotland*

(From Rusk, 78, p. 272)

Area	*Number of Cases*	*Mean IQ*	*SD*
The Four Cities	319	100.86	15.29
Industrial Belt	393	99.19	16.18
Entire Rural Area	162	100.92	14.52
Highlands and Islands	47	101.79	13.13

industrial belt, the entire rural area, and a subdivision of the rural area comprising the Highlands and the Islands, which represent the more isolated rural districts. Not only are there no significant differences between the mean IQ's of any of these groups, but also the highest mean and the smallest variability are found in the Highlands and Islands. In partial explanation of such findings, Rusk, the director of these surveys, observes that "perhaps nowhere has scholastic opportunity been more evenly equated than in Scotland; 99.7% of Scottish teachers are fully trained" [15] (78, p. 273).

It may also be pointed out in this connection that rural living is relatively more desirable and enjoys greater prestige in the British culture than in many other countries. In explanation of the finding that remote rural areas in the British Isles sometimes rate higher on intelligence tests than do those rural areas which are more accessible to cities, it has been suggested that selective migration may have

[15] A similar superiority of the more remote rural areas has been found by Jones, Conrad, and Blanchard (44) in our own New England states. It is noteworthy that in New England educational facilities are probably better, and more nearly uniform from city to country, than in other sections of the United States.

drained the more intelligent persons from the latter areas, but is less likely to have affected the remote rural areas (90). The question of selective migration will be considered in the following section.

SELECTIVE MIGRATION IN RELATION TO URBAN-RURAL DIFFERENCES

Migrations between city and country are constantly occurring for a variety of reasons. During a period of settlement and development, migration occurs predominantly from the urban to the rural districts. The westward expansion of the United States is an example of such a movement. The tide of migration soon turns, however, and the farm dweller is attracted to the city with its promise of wider vocational opportunities and other facilities. At any time, however, such economic events as the opening of mines, the discovery of oil or gold, and to a lesser extent, the construction of roads or the establishment of railway connections will bring about a sudden influx into a previously isolated area. These movements of population, either *en masse* or by single individuals, depend upon a complex manifold of economic, political, social, and psychological factors.

It has been repeatedly argued that the observed intellectual inferiority of rural groups results primarily from selective migration rather from environmental handicap. According to this theory, the more intelligent, progressive, and energetic families or individuals are attracted to urban centers, while the duller and less ambitious remain in the country. The operation of such a selective process for several generations would eventually lead to an inferior rural stock. It is probably true that in certain localities migration may have drained the country of its best families, but this cannot be offered as a universally applicable conclusion. The opposite argument could just as readily be put forth in certain situations, i.e., that it is the shiftless and the dull who migrate because they have been unable to succeed at home. The forces of selection are too difficult to disentangle, unless the specific history and conditions of the district under consideration are known. No single generalization can be applied to all migrations.

The most direct test of the selective migration hypothesis is through a study of the migrants themselves. Such a procedure was followed in a number of studies by Klineberg (49, 50). In one of these, 12-

year-old Negro school boys in three large southern cities were given the National Intelligence Test (49).[16] Those who had migrated to these cities with their families were classified according to length of residence in the urban environment. A city-born group was also tested for comparative purposes. As can be seen from Table 68, a definite improvement in mean National Intelligence Test scores was

TABLE 68 *Mean National Intelligence Test Scores of Southern Negro School Boys in Relation to Length of Urban Residence*

(From Klineberg, 49, p. 54)

Years of Urban Residence	*Number of Cases*	*Mean Score*
One	39	38.3
Two	25	43.2
Three	36	44.7
Four	47	62.5
Five	52	56.2
Six	53	62.2
Seven and more	165	68.7
City-born	359	74.6

found with increasing length of urban residence. The difference is particularly striking if we compare those who had lived in the city for only one year with those who had lived in it for seven years or more. The city-born children, who had been exposed to the urban environment for twelve years, received a still higher mean score. These intellectual differences among the various residence groups may be attributed not only to the varying amounts of time which the subject had spent in a more favorable environment, but also to the age at which such environmental influences operated. Thus since all subjects were 12 years of age, those in the one-year residence group had not been exposed to the urban environment until the age of 11, when they were relatively more "immune" to the effects of environmental changes. The migrant group with the longest urban residence, on the other hand, had moved to the city at the age of 5 or younger. It might be added that there was no apparent basis in economic, social, or

[16] This investigation was part of the general study of selective migration among Negroes which was reported in Chapter 22. In the data now under consideration, however, the problem of northern and southern Negroes does not enter, since all the migrations occurred from rural to urban areas *within* the southern states.

other conditions for any progressive decline in the quality of the migrating population during the period under investigation. Hence the obtained differences in ability between the different residence groups are not likely to have resulted from temporal changes in the nature of the rural groups which migrated to the city.

Another study by Klineberg (50) dealt with white migrants from rural New Jersey to urban centers in the same general area. It was possible to examine the records of 597 migrant children who had taken intelligence tests in rural schools prior to their urban migration. These children were found to average slightly *below* the nonmigrants in the same rural schools. The results of both of these studies suggest that the migrating populations did not represent an initially superior selection, but that they gradually improved *after* moving to the superior urban environment.

It should be noted that both of these studies were concerned with *children*, who did not themselves initiate the migration but simply moved with their families. Somewhat different results have been reported in studies on *adult migrants*. In such cases, the individuals studied are usually the ones who made the decision to migrate. In this respect, these studies might be said to be more direct. At the same time, it might be noted that from the viewpoint of long-range effects over a period of several generations, the data on the children of migrants are actually more relevant.

In general, the studies on adult migrants do show a tendency for the migrants from rural to urban areas to constitute a superior sampling of the rural population (27, 62, 80). In the most extensive of these studies, Gist and Clark (27) followed up a sample of 2544 high school students in forty rural communities in Kansas. All these students had taken the Terman Group Tests of Intelligence in 1922–23, when their median age was 16 years. In 1935, when the median age of the group was 29, the investigators obtained data on the residence of these former students. Over 70% of the original group were found to have migrated, the investigators pointing out that the proportion would be even greater if they had included those who had left the area and could not be located. Of those actually found, 37.89% were living in urban communities, 32.19% had moved to other rural areas, and 29.92% had remained in the original communities. Several comparisons among various migrating and nonmigrating groups revealed statistically significant differences in *initial*

IQ. Thus the migrants to urban centers excelled both the non-migrants and the migrants to other rural areas; those who had moved to larger, cosmopolitan centers surpassed those who had moved to smaller cities; those who had left the state averaged significantly higher than those who had remained in Kansas; and rural non-farmers scored significantly higher than farmers.

Two points should be borne in mind in interpreting these results. First, since the original sampling consisted of high school students, no information is provided regarding the lower levels of the population. There is some evidence to suggest that migrants may be drawn from the extremes of the distribution (26, 102). Thus among moderately successful persons, it may well be that the more alert, ambitious, and intelligent are attracted by the superior opportunities offered by the cities. But among those who are eking out a bare existence close to a subsistence margin, it may be the more hopeless and destitute who are more likely to migrate. A second point is that selective migration does not imply a hereditary interpretation of urban-rural differences. If it should be demonstrated conclusively that the superior families tend to migrate to cities, such families may be superior because of environmental factors within their original surroundings, and their offspring may in turn be superior because they are reared in a relatively favorable family milieu.[17]

SPECIFICITY OF INTELLECTUAL DIFFERENCES AMONG SOCIO-ECONOMIC GROUPS

There is a growing tendency to envisage group differences in terms of specific abilities rather than in terms of general intellectual inferiority or superiority. The application of this concept to racial and national comparisons has already been discussed (cf. Ch. 21). In that connection, it was pointed out that each culture "selects" and stimulates certain abilities, skills, and fields of knowledge as the most significant. Through the fostering of certain talents, specific patterns of psychological development may be produced within each culture. Under such conditions, any attempts to evaluate the mentality of one culture in terms of another would be misleading and would tend to give a decided advantage to the group within which the measuring instrument was standardized. The same may be true of urban-rural

[17] Cf. discussion of similar point in Chapter 22.

comparisons. Intelligence tests have been standardized predominantly on city children, because of the greater accessibility of the latter in large numbers. Consequently, such tests may be overweighted with items which favor the city child, and may fail to sample adequately those abilities in which the rural child excels.

There is a certain amount of evidence which supports such an interpretation of *urban-rural differences* in intelligence test performance. Rural children are neither uniformly nor consistently inferior to urban children on all tests. Thus the city child may excel on the conventional tests of abstract intelligence, but the country child may excel on tests of mechanical or musical aptitude (cf., e.g., 81). Moreover, an analysis of performance on different *intelligence test items* suggests that the relative difficulty of individual items may vary considerably for urban and rural populations. This was clearly brought out in a study by Jones, Conrad, and Blanchard (44). The subjects were 351 children between the ages of 4 and 14, all living in rural areas of Massachusetts and Vermont. The IQ's of these children on the 1916 Stanford-Binet were consistently inferior to the test norms, which had been obtained on a predominantly urban sampling.

An examination of the performance of the rural children on each test of the Stanford-Binet scale, however, showed that these children were significantly inferior on only six tests. Their inferiority on other tests, although consistent from age to age, was statistically insignificant. The tests which yielded the largest degree of rural inferiority were: those involving the use of paper and pencil, as in copying a square; those depending upon specific experiences more common in an urban environment, such as familiarity with coins, street-cars, etc.; and distinctly verbal tests, such as vocabulary and the definition of abstract terms. In all these cases, the *specific environmental handicap* of the rural child is apparent.[18] Similarly, "growth curves" for individual tests showed the greatest age divergence between urban and rural groups in such tests as vocabulary, dissected sentences, naming sixty words, and word definitions. A diminishing urban-rural difference with age, on the other hand, was found in such tests as ball-and-field,

[18] The influence of environmental differences is further demonstrated by the subjects' performance on the four Pintner-Paterson tests administered in this study. On the Mare and Foal, the rural children surpassed the norms, their mean IQ being 110 (cf. similar results obtained with this test by Baldwin, Fillmore, and Hadley, as reported on p. 817). On the Five-Figure Form Board and the Knox Cube, involving more abstract materials, they were slightly inferior; and in Digit-Symbol Substitution, a paper-and-pencil test, they were very inferior.

giving the number of fingers on the two hands, counting thirteen pennies, and other predominantly non-verbal tests.

The investigators also demonstrated that the relative difficulty of individual tests in the Stanford-Binet, as determined by the per cent of children passing each test, may differ for urban and rural groups. For the rural children in their sampling, tests located within a single year level were often quite dissimilar in difficulty. In fact, the range of difficulty within a single year level was sometimes greater than the difference between successive year levels

Such results suggest that the selection and placement of test items might be quite different if a test were standardized on rural rather than on urban samplings. A direct attack upon this question is to be found in a study by Shimberg (83). The basic plan of this investigation was to standardize one test on city children and a second, parallel test on country children. Both tests were then administered to both urban and rural groups, and the performance of the two groups on *each* test was compared. The particular test selected for this purpose was an information test. This choice was justified on the grounds that, in the first place, such tests are frequently included in intelligence scales. Secondly, even in scales which do not contain a separate information test, specific items of information are required in nearly all other tests. Thus a picture completion test, for example, implies the possession of information regarding the characteristic appearance and function of presumably familiar objects.

Each form of the information test consisted of 25 questions. The tests were "scaled," i.e., the questions were arranged in order of difficulty and represented approximately equal increments of difficulty from the easiest to the hardest. This was accomplished by giving a large number of questions to the standardization groups and tabulating the percentage of children who answered each correctly. From these percentages, the difficulty value of each question was computed in terms of σ-units. In the final step, the 25 questions which were most evenly spaced in difficulty value were selected for inclusion in the *scaled test*. This procedure was followed with 764 urban children for Information Test A and with 416 rural children for Information Test B. It should be noted that no question dealing with items of purely local knowledge was included in either form. Both forms were "fair" to city and country children in the sense that the subjects in either group had some opportunity to acquire the requisite informa-

tion. There were, in fact, a number of items common to the two forms. In the original series from which the scaled items were selected, 37 questions were identical in forms A and B.

Both scaled tests were administered to two new groups of urban and rural children. The number of subjects employed in this part of the study was distributed as follows:

	Urban	*Rural*
Form A	6477	610
Form B	902	4875

The mean scores of urban and rural samplings on forms A and B of the test [19] are shown in Figures 100 and 101, respectively.

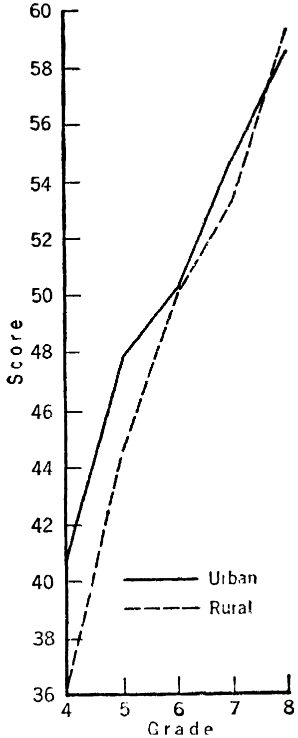


Fig. 100. Average Scores of Urban and Rural Children on Information Test A, Scaled on an Urban Sampling. (Data from Shimberg, 83, p. 45.)

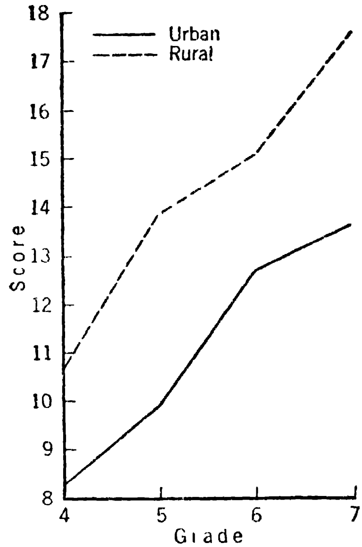


Fig. 101. Average Scores of Urban and Rural Children on Information Test B, Scaled on a Rural Sampling. (Data from Shimberg, 83, p. 50.)

[19] The scores on Tests A and B are not expressed in the same terms. The former are transmuted T-scores with a mean of 50; the latter represent the actual number of correct items out of 25.

It will be noted that on Test A, which was constructed and scaled on city children, the urban groups excel (Figure 100). Among fourth grade children, this difference is 6.5 times as large as its standard error, and among fifth-graders 4.5 times as large. In the upper grades, the critical ratios are under 3, and at the eighth grade the difference is reversed, indicating a very slight and insignificant superiority of the rural group. This reversal is attributed by Shimberg to the differential operation of selective factors in urban and rural groups at the upper school grades. Rural children as a whole, and one-room school pupils in particular, tend to be more retarded educationally than urban groups. Consequently, a large percentage of the duller children have left school before reaching the upper grades, and those who remain are a relatively select group. It might be added in confirmation of this explanation that *age comparisons*, from 9 to 16, revealed a consistent superiority of the urban groups on Test A. In terms of age, the rural children were approximately one year retarded on this test.

On Information Test B the situation is entirely reversed (Figure 101). The rural group is now consistently superior, the differences in its favor being completely significant at each grade. The critical ratios of these differences are all over 3, ranging from 5.56 to 9.33. Thus the hypothesis which this investigation undertook to test seems to have been completely verified. The urban group excelled on the test constructed on city children, the rural group on that constructed on country children. Either group might be ranked "superior," depending upon the specific test employed.

More recent investigations have provided preliminary data which suggest that a similar specificity of intellectual differences may characterize the abilities of different *social classes*. Such a finding was already suggested in an earlier section by the comparison of children of high and low socio-economic levels on different types of intelligence tests. It will be recalled that the inferiority of the lower status classes tended to be greater on the more highly verbal tests. The analysis of intelligence test *items* shows even more marked specificity. This was illustrated in a comparison of the Stanford-Binet performance of 140 first grade children of low socio-economic level with that of 114 first grade children of high socio-economic level (79). The former group did significantly better than the latter on tests involving counting, the handling of money, and sensory discrimination; the latter excelled significantly on tests involving vocabulary, verbal com-

position, rote memory, naming similarities and differences, and motor control. The Stanford-Binet as a whole also showed a larger difference between these two groups than did the Goodenough Draw-a-Man Test.

Of special interest in this connection is the research undertaken by Haggard, Davis, and Havighurst (18, 33) on "cultural differentials" in intelligence test items. The first object of this project was to measure the relative success of children of different socio-economic levels on the individual items of eight widely used group tests of intelligence. The data were obtained by administering these tests to all children of ages 9, 10, 13, and 14 in "Midwest," the same community investigated in the studies by Davis and Havighurst which were discussed in an earlier section. The tests were found to vary widely in the proportion of their items which favored children of high socio-economic status. Within any one test, wide differences in this respect were also found from one item to another. For example, an item based upon an understanding of the term "sonata" was passed by 74.2% of children in a high socio-economic group and only 28.5% of those in a low socio-economic group, whereas an item involving the classification of cutting tools was passed by 76% of the high and 79% of the low socio-economic group.

These differences in the proportion of children of high and low socio-economic status who pass an item are what the investigators mean by the "cultural differential" of the item. They maintain that items with high cultural differentials should be eliminated from intelligence tests, in much the same way that items which favor either sex to a marked degree are now commonly eliminated. Whether or not such items should be included in a psychological test depends, of course, upon the purpose for which the test is designed. If it is our aim to study the ways in which social status classes differ, then it is just the items which differentiate maximally between such classes that are of most interest.

In an effort to analyze further the basis of cultural differentials in intelligence test items, the same investigators (18, 33) set up a carefully controlled experiment which they conducted on 656 11- to 12-year-old children. The children were divided into a high and a low socio-economic group, the two groups being matched in age, school grade, and IQ. The effects of the following factors were investigated: oral *versus* written presentation of items; the interpolation

of two hours of instruction and practice with problems similar to those on the tests; motivation, in the form of a promised movie ticket, for good work during the practice period or during the test proper; and the substitution of test items similar to those in the original test, but rendered "culturally fairer" in their specific content. The lower-class children tended to profit more than the upper-class from the last-mentioned revision of items, although both groups did somewhat better on the "culturally fairer" forms of the tests. Lower-class children also gained relatively more from oral presentation and from added motivation during the test proper.

It is apparent from all these studies that urban-rural, occupational, and other socio-economic groups differ in *specific* ways. Hence any statements regarding intellectual "inferiority" or "superiority" of such groups need to be qualified fully as much as do comparisons between the broader cultural groupings discussed in earlier chapters. Similar qualifications would of course apply to any generalizations regarding emotional adjustment or other personality characteristics, in which group differences are likewise rather specific.

CHAPTER

24

The Individual as a Member of Multiple Groups

DIFFERENTIAL PSYCHOLOGY, in its broadest sense, is concerned with all variations in behavior phenomena among individuals and among groups. The observation and measurement of such differences have led to the accumulation of a vast body of descriptive material which has proved scientifically interesting and practically useful. Examples of such material have been given throughout the present book. The fundamental aim of differential psychology is not, however, the collection of descriptive material. Its aim is similar to that of all psychology, viz., the *understanding of behavior.* Differential psychology approaches this problem through a comparative analysis of behavior under varying environmental and biological conditions. By relating the observed differences in behavior to other known concomitant phenomena, it should be possible to tease out the relative contribution of different factors to behavioral development. If we can determine why one person reacts differently from another, we shall know what makes people react as they do.

The unit of differential psychology is the individual, conceived as a reacting organism; our interest in groups is only secondary. Many traditional groupings, furthermore, have proved to be arbitrary and ill-defined. From the standpoint of behavioral development, the effective groupings are stimulational and not biological. It is not the race, or sex, or physical "type" to which the individual belongs by heredity that determines his psychological make-up, but the cultural group in which he was reared, the traditions, attitudes, and points of view impressed upon him, and the type of abilities fostered and encouraged. Even when behavioral differences are found to be associated with physically defined groups, it is usually the indirect social effects of such groupings, rather than their biological characteristics, which influence behavior development.

Since all types of behavior are influenced by the subject's stimulational background, it follows that psychological data obtained within any one cultural group cannot be generalized to cover all human behavior. Many statements offered under the heading of general psychology are not general at all, but are based upon human behavior as it develops within a single culture (cf. 11, 26, 49). This limitation has sometimes been described as a "community-centrism" which pervades much of our psychological information (49). It has been suggested (26, p. 256) that many textbooks of "general psychology" might be more accurately characterized as dealing with "the psychology of Americans and Western Europeans of the late eighteenth and early nineteenth centuries." In a somewhat similar vein, Dollard (10, p. 17) stresses the importance of considering the cultural setting in which behavioral observations are made. He ventures to suggest that "to the social psychologist, the three most indispensable letters in the alphabet are I.O.C. (in our culture)," and points out that these qualifying letters should be regarded as implicit in all descriptions of behavior within our cultural setting. Such cultural restrictions undoubtedly apply to much of the *descriptive* and *factual* content of psychology. This does not, however, preclude the possibility that when the specific behavior is studied against the individual's stimulational background, the same *principles* of behavior will be found to operate (cf. 12, 14). Such a study of group and individual differences in behavior should, in fact, help to clarify the common underlying principles of behavior development.

CULTURAL FRAMES OF REFERENCE IN BEHAVIOR

The observations of anthropologists in various cultures provide innumerable illustrations of the influence of cultural "frames of reference" upon behavior.[1] What is often regarded as a "natural" response to a particular stimulus may be "natural" only because of the social norms and standards which we have acquired in our own cultural setting. Our very conception of the world about us is influenced by our own specific reactional history. A purely "impartial" or "objective" observer is a psychological impossibility. Each individual's observation and description of any fact is conditioned by his

[1] For a fuller discussion of this point and many additional illustrations, cf. Klineberg (26, 27) and Sherif (49, 50).

special past experiences as well as by the more general traditions and customs inculcated by his group.[2]

Even the simplest *perceptual responses* show evidence of such a cultural framework. Whether we perceive an object as light or heavy, long or short, hot or cold, pleasant or unpleasant may depend in part upon our previous, socially determined experiences. An interesting example is provided by the perception of family resemblances and differences in a number of primitive cultures. Malinowski (32) reports, for example, that among the Trobrianders, resemblance to the father is considered natural and proper, whereas the child is never said to resemble the mother or any of the maternal relatives. The existence of the latter types of resemblance is vigorously disclaimed. Resemblances between brothers are likewise denied, although the resemblance of each brother to the father is granted! It is, of course, difficult to determine to what extent these reactions represent a refusal to admit the proscribed resemblance, and to what extent they indicate a failure to perceive the similarities of appearance. The results of many experiments on the effects of expectation and "set" upon perception, however, make it appear entirely plausible that the Trobrianders only notice those familial resemblances which have been institutionalized by their culture.

It is well known that preferences for tastes and odors, as well as likes and dislikes for foods, vary widely from one culture to another. Among certain African tribes, cologne and scented soap evoked loathing and disgust (cf. 27, p. 209). On the other hand, odors which we find very unpleasant have at other times or places been used as perfumes.

Popular conceptions of time and space, although commonly taken for granted, can readily be shown to be culturally determined (20, 27, 49, 53). Even the concern with precise estimation of the time of occurrence and duration of events, so characteristic of our culture, is quite lacking in others. The indifference to time found among many primitive groups is illustrated by individuals' lack of knowledge about their own age, and by their inability to indicate how long ago an event occurred if a period of several years has intervened. Other differences are to be found in the way in which time is reckoned.

[2] From this point of view, one may regard the instruments and techniques of science as a means of reducing or minimizing the effect of the observer's idiosyncrasies.

The use of astronomical events as a framework for the measurement of time is by no means universal, many other familiar and recurrent events serving for this purpose among different peoples. Thus in Madagascar the natives refer to "a rice-cooking" when they wish to indicate an interval of about an hour, and to "the frying of a locust" to designate a much shorter lapse of time. In the Andaman Islands, it is possible to identify a succession of characteristic odors during the year, as different plants come to bloom. Odors also play an important part in the magic of the Andamanese. It is not surprising, therefore, to find that these people have "adopted an original method of marking the different periods of the year by means of the odoriferous flowers that are in bloom at different times. Their calendar is a calendar of scents" (46, p. 311).

In the Turkish village of Karlik (50, pp. 378-379), which until recently was relatively untouched by modern technological developments, times of the day are indicated by such expressions as "first rooster," "leaving of oxen" (for grazing), "mid-morning," and "return of oxen," rather than in terms of hours. Few of these villagers know the names of the days of the week, or calendar dates and months. Week days are distinguished largely as market days in various neighboring towns, where the villagers must go for all trading. Thus Sunday is *Sandliki* Market, Monday is *Garesar* Market, and Thursday is *Çal* Market. Divisions of the months are based upon the appearance of the moon. In dividing the year, such seasons as summer, fall, winter, and spring are recognized, and to them are added the seasons for "haying," "end of harvest," "sowing," and other farming and animal-raising activities.

Space concepts are equally dependent upon culturally determined frames of reference. That individuals' conceptions of distance and geography are largely colored by their own experiences is at the basis of the waggish "maps" of the United States which have been prepared to portray, for example, a Bostonian's or a New Yorker's idea of the country. Like most caricatures, these "maps" undoubtedly reflect some bona fide perceptual differences resulting from the different interests, traditions, and knowledge of persons reared in different parts of the country. Other examples could easily be cited from everyday observation. The seasoned air traveler has a very different conception of distances than does the farmer who has never ventured farther than the village store. In the previously mentioned Turkish

village of Karlik (50), distances are not ordinarily reported in kilometers or some other standardized system, but are described in such terms as "within a bullet's reach," or "as far as my voice can go." More remote points are indicated in terms of the length of time required to reach them on foot. Confusions and misconceptions arise on the few occasions when such persons travel by train or bus, since they have no basis for translating the time spent in transit into their familiar frame of reference. A similar use of such "psychological units" to express distance has been observed among the Saulteaux Indians, who estimate distance in terms of the number of "sleeps," or nights spent on the road (21).

Even the familiar designation of directions in terms of north, south, east, and west, although prevailing over a large part of the world, is not a universal system. Thus among the natives of Dobu, space is conceived as a large garden clearing, such as the individual encounters in the daily life of his community. "Just as the garden has its inland border *kaikai*, its seaward border *kunnkunnwana*, and its sides *nana*, so also has space in its widest extension" (17, p. 131).

The individual's *memory* for events he has observed or facts he has been told is likewise colored by his cultural background. This is particularly well illustrated by the observations and tests of Bartlett (3, 4) and Nadel (40, 41). Both investigators, working with South African tribes, have shown the important part played by cultural patterns in the "restructuralization" and distortions of recall. For example, when repeating a European story, individuals in these groups tended to cast it with characters typical of their tribal folklore. They likewise rearranged sequences and introduced twists of plot characteristic of their native stories.

Bartlett (3) reports a number of interesting observations made among the Swazis, a small group of Bantu natives of South Africa. In recounting an episode, or in answering a simple question, the Swazis characteristically repeat every detail, however irrelevant, and appear incapable of reaching the end of the story without a rote recital of all the intervening steps. An especially vivid illustration of the effect of socially determined set upon recall is provided by an incident cited by Bartlett. A legal matter necessitated a visit to England by a Swazi chief and several of his leading men. When the party was questioned upon its return, their most vivid recollection was that of the English policeman regulating traffic with uplifted hand. In

explanation, Bartlett writes: "The Swazi greets his fellow, or his visitor, with uplifted hand. Here was the familiar gesture, warm with friendliness in a foreign country, and at the same time arresting in its consequences. It was one of the few things they saw that fitted immediately into their own well-established social framework, and so it produced a quick impression and a lasting effect" (3, p. 248).

Studies by the "*free association*" technique provide further indications of the effects of conventionalized stimulational backgrounds. In one of the most clear-cut of such experiments (16), 218 male subjects, including first and second year law students, first and second year medical students, and liberal arts students, were tested with a forty-word list. The list contained twenty "neutral" and twenty "critical" words arranged in random order. The critical stimulus-words were chosen so that each had a common legal meaning, a different medical meaning, and if possible a third neutral meaning not specifically related to either field. The neutral stimulus-words were selected from an earlier standardized list (Kent-Rosanoff); they were words to which people in general gave relatively uniform responses which were neither legal nor medical in nature. Examples of the neutral stimulus-words included: rough, girl, long, river, eagle. Some of the critical words were: hereditary, expiration, discharge, compensate, void, tender.

A classification of the responses to the critical words showed that the first and second year law students gave 8% and 17% more "legal" responses, respectively, than the control group; and they made 4% and 5% fewer "medical" responses, respectively, than the control subjects. The first and second year medical students, on the other hand, made 25% and 30% more "medical" responses, respectively, than the control group; their corresponding frequencies of "legal" responses were 9% and 11% fewer than those of the control group. Typical "legal" responses to one of the critical stimulus-words, "administer," included such terms as administrator, estate, government, money, will; among the "medical" responses to the same stimulus-word were anaesthetic, dose, first aid, inject, syringe. The results of this experiment demonstrate the influence of occupational conditioning upon verbal reactions. Similar differences would undoubtedly be found with reference to broader cultural groupings.

Cultural influences are also discernible in the *motor habits* of different peoples. The gait and tempo of walking, as well as the charac-

teristic standing, sitting, and sleeping postures, vary widely from one culture to another. The carved ivory and wooden headrests of Africa which are preserved in our art museums impress the American observer as a most uncomfortable sleeping aid! Most primitive peoples sit in a squatting posture; Eskimos, as well as many American Indian groups, habitually sit on their heels (cf. 6).

The role of cultural factors in the development of gestures has already been discussed in an earlier chapter (Ch. 22). A typical illustration of the cultural conditioning of a response often assumed to be "natural" and universal is to be found in gestures of negation and affirmation. Nodding to signify assent is by no means shared by all peoples, nor is the lateral turning of the head a universal sign of negation. The Semang, a pygmy tribe of interior Malaya, say "yes" by thrusting the head sharply forward, and "no" by lowering the eyes (30). The Dyaks of Borneo raise their eyebrows for "yes," and contract them slightly for "no." For the Maori, raising the head and chin signifies "yes"; among the Sicilians, the same gesture means "no" (26, p. 282). The use of the fingers in pointing is likewise restricted to certain cultures. Among several American Indian groups, for example, pointing is executed with the lips (30). Changes in gesture patterns occurring over a relatively short period within our own culture can be readily noted by looking at early movies. The seemingly "unnatural," stilted, and exaggerated nature of the actors' gestures is immediately evident to the modern observer.

Closely related to such observations of gesture is the comparative study of *emotional expression* in different cultures. A rich body of data is available in this area, indicating differences in the extent of emotional display, the occasions on which emotional behavior is manifested, the specific patterning of emotional responses, and the degree of control which the individual is able to exert over such behavior (cf. 27, Ch. VII; 30). The manner of greeting employed by different peoples would in itself constitute a fertile field for such inter-cultural comparisons. The practice of kissing, as a friendly greeting or as a sexual response, varies widely in different cultures and is totally absent in a number of primitive societies. It is interesting to note in this connection that Kinsey and his associates (28) found similar differences between social classes within our own culture. Among persons of lower socio-economic level, kissing as a sexual response was relatively infrequent and was even regarded by many as unhygienic. The latter

attitude was particularly interesting in view of the fact that it was frequently expressed by persons who habitually used common drinking cups and followed other practices considered insanitary in higher socio-economic levels.

Many instances have been recorded of the ceremonial control of certain emotional reactions, such as weeping, to a degree which appears surprising to persons reared in our culture. Ritual shedding of tears on a variety of prescribed occasions has been observed in such countries as China and Montenegro, in a number of American Indian tribes, and among the Maori, Andaman Islanders, and other primitive cultures (27, Ch. VII).

An especially clear illustration of the effect of the cultural milieu upon behavior is furnished by *aesthetic preferences* and artistic "taste." The evolution of styles in music, painting, sculpture, architecture, and the other arts testifies to the shifting demands of "taste." The styles which are derided as harsh, barbaric, and uncouth by one generation have often been accepted as masterpieces by the next. Any artistic innovation which clashes too vigorously with the familiar and the traditional forms of artistic expression requires a period of gradual habituation. It is an unfortunate but perhaps psychologically indispensable fact that the great art leaders who are subsequently hailed as the initiators of new movements often suffer ridicule and derision during their lifetime. This follows necessarily from the fact that they come at a time when the adequate experiential background for the enjoyment of their products is lacking.

The question of the sophisticated and the naïve observer is also relevant to this point. The trained critic or the sophisticated observer has had certain specific experiences which enable him to enjoy artistic products that may appear meaningless, indifferent, or even unpleasant to others. Psychologically, there is no "naïve observer"; such an individual is naïve only from the standpoint of a specific class of experiences. His judgments are, however, directly influenced by other experiences which he has had. His artistic reactions will be largely dictated by common everyday observations and popular fashions. Thus the observer may enjoy realistic art because he is more familiar with photographic reproductions of objects; or he may reflect some traditional artistic conception which has been inculcated in him from early childhood. But in no case is his judgment made independently of experience. The essential difference between the sophisticated and

the naïve observer is in the *kind* of past experience which they have had.

It is a familiar observation that Occidentals who hear Chinese music for the first time find it not only discordant and harsh, but also unpleasantly loud. At the same time, it has been reported that the Chinese find American jazz and Wagnerian brasses disturbingly loud upon their first exposure to such music (cf. 27, p. 209). Similar data are provided by the history of Western music, which clearly reveals a progressive shift in the point of demarcation between consonance and dissonance (cf. 39). Intervals which were considered dissonant at one period were accepted as consonant in the next. The transition occurred from those intervals in which fusion of the notes is easily obtained to those in which fusion is more difficult. As the newer intervals came into use, the intervals which fused more readily declined in popularity. The preferred intervals at any one period seem to have been those which were "just consonant," i.e., those in which fusion was neither too easy nor too difficult. The former were regarded as relatively uninteresting, the latter as dissonant.

As an experimental check upon such a "genetic" theory of consonance, Moore (39) analyzed the repeated judgments of nine subjects on four musical intervals. Two of these intervals were considered dissonances (major and minor 7th) and two consonances (3rd and 5th) at the beginning of the experiment. The subjects underwent a period of habituation in which all four intervals were repeatedly experienced in musical passages. The judgments obtained at the end of this period showed certain unmistakable changes in the relative preference for each interval. Of the two initially consonant intervals, the 3rd lost rapidly in aesthetic value, while the 5th maintained a fairly constant level. The dissonances, on the other hand, showed a gain in preference, the minor 7th gaining more rapidly than the major 7th. According to Moore's theory, the region of highest aesthetic value for an interval is the "barely consonant" region. The greatest changes with repetition were therefore to be expected in the intervals nearest this region, viz., the 3rd on the one hand and the minor 7th on the other. This experiment furnishes a vivid demonstration of the dependence of artistic "taste" and aesthetic judgments upon experiential factors.

Another relevant experiment is that conducted by Foley (13, 15) on occupational differences in preferential auditory tempo. The subjects were 684 girls between the ages of 13 and 20, all of whom were

enrolled in a trade school in New York City. Comparisons were made between groups in the following courses: power machine sewing (N = 90); hand sewing (N = 180); beauty culture (N = 165); typewriting (N = 182); and courses in domestic occupations, including waitress training, home nursing, nursery education, and others (N = 67). These vocational groups were roughly comparable in age, intellectual status, education, socio-economic level, and natio-racial background. The principal difference between them was in their specialized vocational training. In the experiment, each subject listened individually to a series of auditory tempos produced with a standard metronome. Six representative speeds were used, ranging from 56 to 200 beats per minute and corresponding to the musical designations of largo, larghetto, adagio, andante, allegro, and presto. The six speeds were presented serially, in an ascending and then a descending order, the subject reporting whether she liked or disliked each. The procedure was repeated, when necessary, until the preference was narrowed down to a single speed in both the ascending and descending series.

TABLE 69 *Auditory Tempos Preferred by Different Vocational Groups*

(Adapted from Foley, 15, p. 125)

Vocational Group	*N*	*Mean Preferred Tempo*	*Approximate Musical Designation*
Typewriting	182	178.08	Allegro-Presto
Power machine sewing	90	161.02	Allegro
Beauty culture	165	139.04	Andante
Hand sewing	180	134.46	Andante
Domestic occupations	67	133.61	Andante

The mean tempos preferred by each vocational group are shown in Table 69. It is apparent that wide occupational differences were found, which reflected the nature of the auditory stimulation to which the subjects had been exposed during their vocational training. Thus the beauty culture, hand sewing, and domestic groups chose the slower rates, corresponding closely to an andante tempo. The power machine sewing group preferred a relatively slow allegro, while the typists chose a fast allegro bordering on presto. Although individual differences within each group were large, the mean differences showed a high degree of statistical significance. Moreover, the group differences

in preferential tempo were more marked between advanced groups than between groups which were at an earlier stage of vocational training. The findings of this study are in close conformity to what would be expected on the basis of occupational conditioning. The machine sewers and typists, who preferred the more rapid tempos, had been exposed to loud, rapid, repetitive noises from the typewriters and power sewing machines. On the other hand, rapid auditory stimulation does not accompany the activities performed by the hand sewing, beauty culture, and domestic groups, which preferred the slower tempos. The habitual occupational activities thus seem to have shifted the subjects' frames of reference, in terms of which their preferential judgments were made.[3]

We need not go beyond everyday observations in our own culture to find further evidence of shifting frames of reference in preferential responses. A vivid illustration is provided by the response to changing fashions in women's wear. A style which appears beautiful to most observers when it is at the height of fashion will probably look dull and unattractive within a season, and positively ludicrous if viewed ten years later. These rapid changes in "taste" come as no surprise to fashion leaders, since the fashion industry deliberately provides the stimulation which brings about the change in response. Upon the introduction of a new style, the buying public is exposed to a carefully planned and coordinated "blitzkrieg," designed to prepare them for the acceptance of such a style. The new fashion is pictured in magazines and newspapers; models wear it on the street and in theatres, restaurants, and other public places; window displays feature it conspicuously. Through these and similar techniques, the public is rapidly "sensitized" to the new style in much the same manner that Moore's subjects were habituated to the unfamiliar, discordant combinations of notes. If the fashion is too "discordant" and clashes too violently with the previous experience of the public, the sensitizing process may fail and the fashion will be rejected. Similarly, the successful fashion leader keeps well posted on current developments in other areas—social, economic, political, artistic—in order to coordinate his innovations with the more general frame of reference of his consumer public.

[3] The reader is referred to the original study (13) for a summary of evidence indicating that these group differences are *not* likely to have resulted from selective factors.

Even in the realm of *science,* where "objective truth" presumably reigns supreme, the effect of the individual's frame of reference cannot be wholly eliminated. The data of science admit of various interpretations. One or another of such interpretations may seem to follow inevitably from the given facts, depending upon the observer's experiential background. This is exemplified by the various approaches of different sciences to the same phenomenon, as well as by the presence of distinct "schools" within a single science. There are "fashions" in science as in other areas. The general cultural milieu of the period is reflected in the nature of its scientific products and theories, just as it is in other phases of human activity. It is not a coincidence that certain basic similarities can be found in such diverse phenomena as the science, art, social structure, and economic policies of any given period. The setting for all such developments is the common experiential background of the people of that age.

"DEVELOPMENTAL STAGES" AND THE CULTURAL SETTING

Theories of developmental stages furnish numerous illustrations of the tendency to overgeneralize from observations within a single group. Child psychology is replete with such theories. Much interesting material has been gathered, for example, on the formation of *concepts* in childhood. The child's ideas about the physical world, his "consciousness of self," his interpretation of dreams, and similar conceptions have been analyzed into definite developmental sequences. Outstanding in this field are the theories of the Swiss psychologist Piaget (43, 44, 45).

In an extensive series of investigations, Piaget arrived at the conclusion that the thinking of the child is *animistic* and that the transition from this initial animism to the adult's conception of the world is made through four major stages. For children between the ages of 4 and 6, everything active is alive. Since children of this age are also anthropocentric, "activity" is regarded as synonymous with usefulness to man. Thus the sun is active because it gives warmth, stones are active because you can throw them. At this first stage, therefore, all objects which are unbroken and in good condition are considered to be alive and "conscious." In the next stage (6–7 years), only movable objects are believed to be alive. In the third stage (8–10 years), life is attributed only to things which can move spontaneously. Thus the sun and

a river are alive, but an automobile is not. In the final stage (11 years on), life is restricted to animals and plants, or sometimes to animals only.

Such stages have been commonly accepted as an inevitable or natural development through which the child must pass. There are, however, numerous factors within the experience of a child in our society which might account for such animistic tendencies. The language which the child is taught encourages him to form an animistic conception of the world. Thus he hears the sun referred to as "he," and the moon or a ship as "she." Figurative expressions, such as the "rising" and "setting" of celestial bodies, the "running" brook, and the "howling" wind, are not conducive to an impersonal conception of natural phenomena. If to this are added the fancies of poetry, fairy tales, and other imaginative literature, it is apparent that the child's experience has a strongly animistic flavor. It is not until he has had the opportunity to accumulate a certain amount of information from direct observation of cause and effect in everyday situations, that such a child can arrive at a realistic notion of the world.

Data supporting such an experiential interpretation of the development of children's concepts are to be found in studies on children in different cultures. Mead's observations on the island of Manus in New Guinea led her to conclude that animism is absent in the thinking of Manus children (37). In both the spontaneous remarks of these children and in their replies to questions, she found evidence of a very realistic conception of natural objects and events. The drifting away of a canoe, for example, was not attributed to malicious intent on the part of the canoe or to other supernatural factors, but to the fact that it was not securely fastened. Such an answer was obtained despite the fact that in her conversation with the child the investigator had attempted to place the blame on the canoe. Mead attributed this realistic attitude to the type of training which such children receive. From early childhood they are forced to make a correct adjustment to the physical demands of their environment. The responsibility for a mishap is never shifted to an inanimate object, as in blaming the log if the child trips over it. If the child hurts himself, he is told that it is the result of his own clumsiness. It is interesting to note that, in certain respects, the adults are more animistic than the children in this culture, since they explain sickness, death, and other misfortunes as the activity of "spirits."

Dennis (9) has argued that Mead's results do not disprove the applicability of Piaget's developmental theory to the thinking of Manus children, since her methods of investigation were not comparable to those of Piaget. Mead's findings, according to Dennis, demonstrate only the absence of tendencies to personify and humanize inanimate objects, rather than a lack of animism among Manus children. Whatever the interpretation, however, these observations do suggest that the characteristics of child thought may vary from one culture to another.

In a subsequent investigation by Dennis (9), designed as a more direct check of Piaget's theories, 98 Hopi children were studied through standardized individual interviews and a group questionnaire. The survey dealt with (1) animism in the more restricted sense, i.e., being alive; (2) the attribution of "consciousness" to things; and (3) "moral realism," as in the explanation that "the bridge fell because the boys crossing it had stolen apples." In all three respects, the Hopi children were far more animistic and less realistic in their replies than white children of the same ages tested in other investigations. Dennis dismisses the possible explanation that differences in "intelligence" might account for the greater animism of the Hopi children, since their performance on the Goodenough Draw-a-Man Test equaled or excelled the white norms. In the light of the known characteristics of the Hopi culture, Dennis concludes: "The explanation must, therefore, be sought in terms of environment, and no doubt in the cultural environment rather than in terms of the physical environment. The differences in social environment between the Hopi child and the white American child are numerous" (9, p. 32). He further points out that the concepts of the Hopi children are similar to those reported by Piaget and others for white children, but that such animistic concepts are retained until a later age among the Hopi.

Emotional development and *personality adjustments* have also been analyzed from the point of view of "stages." The most widely discussed of such stages is probably the period of "storm and stress" characteristic of the adolescent. Almost all writers on child psychology ascribe emotional upheavals, personality changes, conflicts, and maladjustments to this age. There is evidence, however, to show that this is not a universal phenomenon. In certain societies (cf., e.g. 34, 35, 38), the adolescent assumes his altered status, both physical and social,

without emotional disturbance. His tasks are cut out for him by tradition; there are no momentous choices and decisions to be made; no mystery attaches to his position; and no trace of embarrassment is encountered.

There is much in our society, on the other hand, which fosters adolescent maladjustments. Thus the individual is placed in an ambiguous and ill-defined position, being treated neither as a child nor as an adult. Restrictions upon his actions are frequently increased, while at the same time he is expected to be more self-reliant than he had formerly been. Embarrassment and a general atmosphere of mystery are often directly induced by adults through their attitudes, remarks, and actions. In view of the many experiential factors in our society which might lead to adolescent maladjustments, there seems to be no need to posit an innate or physiological basis to the storm and stress of this period, nor to regard such emotional upheaval as a necessary developmental stage.

Another aspect of child behavior to which the concept of developmental stages has been widely applied is *drawing*. Children's drawings have been collected in large numbers and submitted to detailed analyses, in the hope that they might furnish a clue to the child's mentality. The best-known example of such a use of children's drawings is provided by the Goodenough Draw-a-Man Test, with its carefully standardized scoring and extensive age norms. The voluminous literature on children's drawings reveals a widespread belief among psychologists in the existence of definite developmental sequences in drawing behavior. These stages have often been regarded as products of maturational factors and assumed to be independent of specific environmental stimulation. The drawings characteristic of each age level are believed to be distinguishable in subject-matter as well as in many aspects of technique and execution.

Such generalizations in regard to the drawing behavior of children are, however, limited to certain specific groups with a common cultural background. Spontaneous drawings by children of different national and cultural groups have been gathered and described by several investigators.[4] These data bring out very clearly the part played by the child's environment in determining every phase of his drawing behavior. Thus the type of object most frequently drawn at each age

[4] For a survey of much of this literature, cf. 1, 19.

shows a wide variation from one group to another. In the studies on American children (cf., e.g., 1, 19), drawings of the human figure predominate at the younger age levels. That this is not a universal tendency among young children has, however, been repeatedly demonstrated. In a study on Swiss children, for example, the human figure occupied a relatively insignificant position, miscellaneous objects and houses heading the list (25). Representations of people are likewise infrequent or almost completely absent in the drawings by children from several other countries (cf. 1). In general, the subject-matter of children's drawings varies so widely from group to group as to make any attempted universal classification quite meaningless.

Similar differences are apparent in all other aspects of the drawings (cf. 1). Whether the child draws broad panoramic views or scenes at close range, isolated objects or organized pictures, imaginative themes or realistic portrayals seems to depend in large measure upon his specific environmental milieu. In certain groups, the drawings are full of action, in others stationary objects and figures predominate. The organization of the picture likewise differs from one group to another. In some groups, a single unified scene is most often presented, in others a sequence of events, in still others isolated•objects. The degree to which color is employed, as well as the choice of specific hues, usually reflects the influence of both physical environment and social traditions.

The representation of detail brings out some interesting facts. In certain groups, detail is relatively poor, total impressions and broad vistas being emphasized. In others, the minutest details are painstakingly drawn into the picture. An even more significant point, however, is the *specificity* of the details which are represented. Thus among a group of children belonging to a hunting tribe in Siberia, remarkably accurate and naturalistic drawings of reindeer and elk were obtained (48). These drawings were clearly superior to those of the human form or of any other subject executed by the same children. It should be noted that none of these children had had previous experience in drawing. The investigator points out that the sharpened visual perception, manual dexterity, and keen observation fostered in such a hunting tribe probably influenced the accuracy of the drawings, especially when the objects represented were the animals commonly hunted by the tribe.

This tendency to elaborate those details which are specifically observed and which play an important part in the individual's everyday activities, while ignoring other details, is very commonly found in children's drawings. Examples can easily be multiplied (cf. 1). Thus in one survey (1), the drawings by European children usually portrayed vegetation only in a general way, with no attempt to show specific type or variety. The children from many tropical and semitropical countries, on the other hand, often featured fruit trees and dense forests as a major part of the drawing, and included sufficient detail to indicate the particular type of plant pictured. One group of drawings by Hungarian and Czecho-Slovakian children, although crude in other respects, showed minute and carefully executed details in the national folk costumes. In drawings by American Indian and Balinese children, elaborate details occurred in the ceremonial masks and head-dress, in contrast to the paucity of detail in other objects.

Stylized representations and special cultural attributes are also discernible in the drawings by children in certain cultures, becoming increasingly apparent with age (cf. 1, 38). In a study of Hopi children with the Goodenough Draw-a-Man Test, the younger children tended to draw generalized human figures, while approximately one-third of the 10-year-olds drew figures in which special characteristics of the Indian culture could be recognized (8). A similar tendency has been noted among Indian children of the Northwest Coast of Canada. In one such survey (2), the instructions were simply to "draw an animal." The subjects consisted of 159 Indian children between the ages of 5 and 18, all attending an Indian school at Alert Bay. About half of the group belonged to the Kwakiutl, the remainder being divided among five other tribes found in British Columbia. Many of the drawings reflected the habitual activities and interests characteristic of life in an Indian community. Twenty were clearly recognizable as stylized representations executed in the traditional manner of the particular Indian culture. Both the subject matter and technique of these drawings showed the influence of the subjects' institutionalized behavior. Among the animals portrayed were the killer whale, sea lion thunder bird, and mythical double-headed serpent. That a mythical creature should be drawn at all in response to the directions to "draw an animal" is itself an indication of the strength of the cultural influ-

ences. The highly conventionalized technique of these drawings can be seen in Figure 102, which shows samples of the killer whale and sea lion. It should also be noted that such stylized drawings were more

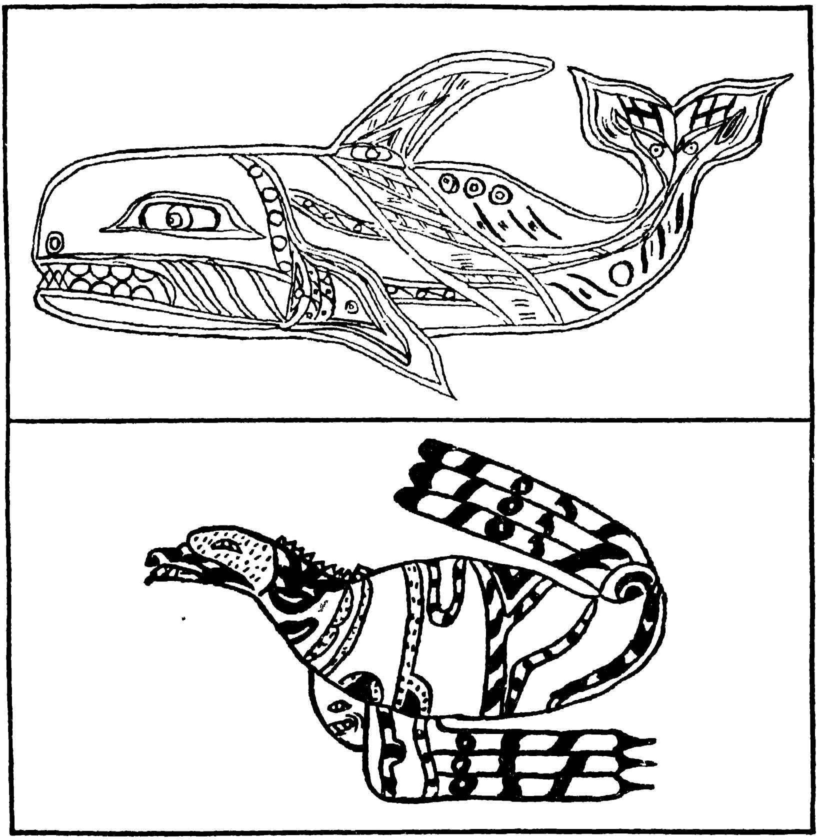


Fig. 102. Drawings by Indian Children of the Northwest Coast, in Response to Directions to Draw an Animal. (From Anastasi and Foley, 2, p. 369.)

common among boys, having been produced by 17% of the boys and only 7% of the girls. This is in keeping with the fact that painting and carving are conducted exclusively by the men in these tribes. In the light of all these findings, it would seem hazardous to regard the richness of detail, general technique, or any other feature of children's

drawings as an index of developmental stage, unless such features are considered in reference to the child's cultural background.

LANGUAGE AS A CULTURAL FACTOR IN BEHAVIOR

The data of comparative linguistics and anthropology—and more recently the writings of the semanticists—have suggested the important part which the nature of a people's language may play in their conceptions of the world about them, their attitudes, and other behavior characteristics (cf., e.g., 7, 22, 29, 31, 33, 42, 47, 54). In a very fundamental sense, language provides the tools for much of our thinking. The relationship between language and thought has been vividly expressed by Whorf (54, p. 231), who points out that each particular language "is not merely a reproducing instrument for voicing ideas but rather is itself the shaper of ideas, the program and guide for the individual's mental activity, for his analysis of impressions, for his synthesis of his mental stock in trade." As he further states, "We dissect nature along lines laid down by our native languages." In a similar vein, Mauthner (33, p. 4) wrote, "If Aristotle had spoken Chinese or Dacotan, he would have had to adopt an entirely different logic, or at any rate an entirely different theory of categories."

Language influences the type of distinctions and discriminations which we make in observing our surroundings. Objects and events in nature do not, of course, occur in the distinct categories to which we have become accustomed. Such categories have generally been developed to fit specific purposes and to facilitate our dealings with objects. Once objects are put into a specific category, or "named," however, our attention is thereby focused upon their similarities or common characteristics, and we tend to ignore differences among members of the class. Thus what we notice and what we overlook in our environment depend in part upon our particular linguistic system. When the conditions existing within a given culture have made certain distinctions important, we are likely to find separate words corresponding to such differentiations.

This is illustrated in Figure 103, in which certain words in the Eskimo and Hopi languages are compared with their English equivalents. Thus to correspond to our one word, "snow," Eskimos have several words, indicating "falling snow," "slushy snow," "hard-packed

snow," and so on. On the other hand, the Hopi use a single word to designate "anything which flies, exclusive of birds." An insect, an airplane, and even a pilot would be called by this single name, the context determining which was meant. For our one word, "water," how-

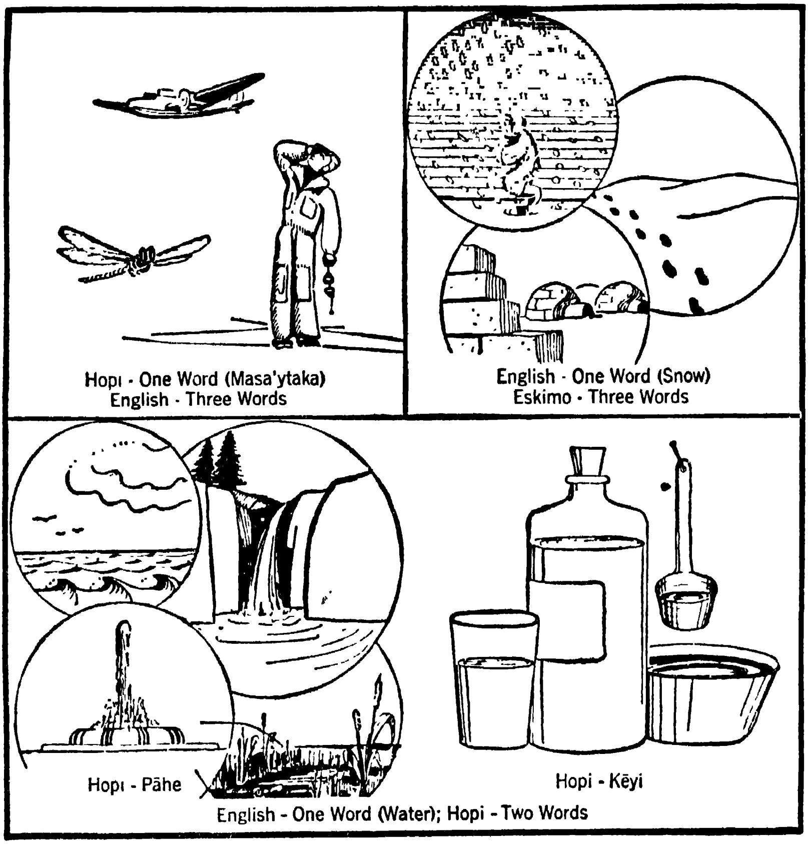


Fig. 103. The Classification of Objects in Different Languages. (From Whorf, 54, p. 230.)

ever, the Hopi have two words, one referring to "flowing water" and the other to "water in one place, held within a container." Examples could easily be multiplied. In Arabic, the number of different words relating to "camel" is said to be about six thousand (51). There are terms to refer to riding camels, milk camels, and slaughter camels;

other terms to indicate the pedigree and geographical origin of the camel; and still others to differentiate camels in different stages of pregnancy and to specify innumerable other characteristics important to a people so dependent upon camels in their daily life.

A particularly interesting example of the role of language in the classification of observed phenomena is provided by *color terminology.* The way in which hues are grouped varies in different languages, and probably in turn affects the type of color discriminations which are customarily made in each particular culture. In certain modern European languages, there are different words for "light blue" and "dark blue," just as in English we have the terms "pink" and "red." The Ashantis of the African Gold Coast have color names for black, red, and white: the term "black" is used for any dark color, such as blue, purple, or brown; while "red" also covers pink, orange, and yellow (52). In the same group, gray is expressed by the word for "wood ashes," and green by the term for "tree" or "leaf" (52). Among the Manus of New Guinea, yellow, olive-green, blue-gray, gray, and lavender are regarded as variations of one color (36). The terms for "blue" and "green" are often combined in primitive languages. We cannot, of course, conclude from these linguistic classifications that the color sensitivity of such peoples is inferior to or different from ours. Objective tests of color-blindness have demonstrated a normal ability for color discrimination, despite the lack of differentiating terminology. It is apparently the specific conditions of the particular culture which determined the type of classification developed in each case.

Not only vocabulary, but also the formal aspects of language, show characteristic differences from one culture to another. Thus in the Hopi language there are no temporal references in verbs, but special forms are employed to indicate the nature of the statement, such as immediate observation, memory or generalization (cf. 54). Similarly, among the Hupa Indians of California, a suffix is used to designate the source of information, such as hearing, sight, or conjecture from circumstantial evidence (18). Our distinction between nouns and verbs is not so fundamental or universal as might be supposed. To take another illustration from the Hopi language, such terms as "lightning," "wave," "meteor," "puff of smoke," and "pulsation" are verbs, as are all events of necessarily brief duration (cf. 54). In some

languages, the distinction between verbs and nouns is non-existent, all terms corresponding most nearly to our verbs. Thus "it burns" would signify a flame, and "a house occurs" or "it houses" would refer to our noun "house" (cf. 54).

In an analysis of the extensive data collected by Malinowski on the Trobriand Islanders, Lee (31) proposed a provocative theory regarding the language of these people and its role in their other behavior. The Trobriand language, according to Lee, shows a focusing of attention upon disparate elements or acts, considered independently, rather than upon relationships among events. Their sentences are composed of essentially unrelated words. The comparative and superlative degree are absent, as are pure adjectival concepts; adjectives in this language refer to specific classes of objects and cannot be abstracted from them. Cause-and-effect relationships are not conventionally expressed. When questioned regarding such causal relations, the individual Trobriander does not have ready-made answers provided by his culture. Each individual must think out his own answer, the replies showing confusion and disagreements. Chronological sequences are likewise unimportant to them. "The past is not an ordered series, but rather a chaotic repository of unrelated events, which, at best, are remembered as anecdotes" (31, p. 360).

A frequently reported observation is the relative scarcity of abstract terms in most primitive languages. Such a condition, of course, makes abstract thought much more difficult. It need not, however, imply an inability to carry on abstract thinking, any more than color-blindness is implied by the color terminologies discussed above. The presence or absence of abstract terms in a particular language may simply reflect the conditions of life within that culture. There is some evidence suggesting that terms of a higher level of abstraction can often be developed by such peoples when a situation is presented which requires such terms (cf. 27, p. 46).

Finally, it should be borne in mind that language is, essentially, *behavior* (cf. 24). It is not an independent entity, as philologists are sometimes inclined to regard it. At the same time, language serves as a potent cultural influence. The particular system of linguistic terms and forms institutionalized by a given culture represents an important part of the total complex of stimulation to which each individual is exposed. Regardless of how such linguistic behavior originally evolved

in the group, it assumes a major role in shaping the psychological development of the individual.

"HUMAN NATURE" IN DIFFERENT CULTURES

Certain ways of acting have long been popularly regarded as "natural." This designation usually implies that the behavior in question is "normal" as well as innate and biologically predetermined. Closely related to this concept are those of "perversion" and "reversion." The former refers to behavior which is considered "unnatural"; the latter implies a revival or reinstatement of a more "primitive" and less "artificial" type of behavior. Thus if one type of behavior is assumed to be *natural,* then any environmentally produced variation of such behavior is considered a *perversion*. Similarly, if a "civilized" person be put in a "primitive" environment, the behavioral changes which may ensue are regarded as a *reversion* to a natural state. The latter is implicitly assumed to have existed all along, but to have been held in abeyance, so to speak, by conditions in a civilized community. It is apparent that the concepts of perversion and reversion have meaning only as long as one specific way of behaving is assumed to be the "natural" way.

It has been repeatedly demonstrated, however, that no one form of behavior is any more natural than another in the sense of being predetermined by innate constitution. The data on this question are derived chiefly from two sources. The first is the *experimental production* of behavioral variations. A number of such experiments on infrahuman organisms have been reported in Chapter 6. The import of their results was to show that different types of behavior will follow as a natural result of varying environmental conditions. Much so-called instinctive behavior has been shown to be natural only under given environmental conditions.

The same point has been demonstrated by *inter-cultural comparisons*. Many forms of behavior which have been labeled "instincts" and "fundamental drives" are found to differ significantly from one cultural group to another.[5] Thus the role of cultural factors in the expression of the *maternal drive* is illustrated by the widespread custom of

[5] For many illustrations of this point, cf. Klineberg (27) Ch. V and VI, and Sherif (49).

adopting children, which is practiced among several Melanesian, South African, and American Indian groups. In certain tribes, children are so infrequently reared by their own parents that it is very difficult to obtain genealogies. Similarly, in ancient China the social concept of maternity was distinct from biological maternity. Thus all offspring of "secondary wives" within the family unit were considered to be children of the "first wife." The latter was the only person in the role of mother, the other wives being indiscriminately regarded as "aunts" by their own as well as other children in the family (cf. 27).

Aggressiveness and *fighting,* popularly considered to be among primitive man's natural impulses, are unknown among several groups. In a few tribes, for example, no weapons or implements of warfare are to be found. That men should attack each other seems inconceivable to individuals reared in such cultural groups. Similarly, *acquisitiveness* and the desire for personal property are not a universal phenomenon.[6] A striking demonstration of this fact is provided by the social institution of the *potlatch,* as found among the Indians of the Canadian Northwest Coast. In this culture, social prestige is achieved through the distribution or giving-away of personal property, rather than through its acquisition.

The manifestations of the *sex drive,* with its attendant feelings such as love and jealousy, likewise exhibit wide inter-cultural variations. The diverse customs and conventions associated with mating behavior in different groups have been extensively described by anthropologists and many are undoubtedly familiar to the reader. It will be recalled that similar differences in the typical manifestations of sex behavior were found by Kinsey *et al.* (28) in their comparisons of different socio-economic classes within our own culture. Mention may also be made of the sets of *rules and restrictions* imposed by different societies upon many forms of behavior, including aggressiveness and physical violence, reaction to personal property, sex activities, and others. The wide variations in such restrictions show them to have little or no basis in "human nature" as such. The *mores* of one society often appear as a quaint set of taboos to another.

The traditional *sex differences* in abilities and in personality traits are another case in point. It was long considered "natural" for the sexes to differ in general intelligence and especially in aptitude for scientific pursuits and similar branches of learning. Men, too, were

[6] Cf., e.g., Beaglehole (5) for a comprehensive treatment of this question.

regarded as naturally more stoical, less given to emotional displays, more competitive, less sympathetic. If a given individual displayed the intellectual or personality traits of the opposite sex, this was considered "unnatural." An understanding of the experiential basis of these behavioral characteristics shows the artificiality of such a distinction between natural and unnatural behavior.

NATURE AND VARIETY OF PSYCHOLOGICAL GROUPS

Psychologically the individual belongs to every group with which he shares behavior.[7] From this point of view, group membership is to be defined in terms of behavioral rather than biological categories. The effective grouping is not based upon the individual's race or sex or body build, but upon his experiential background. Thus if the individual is reared as a member of a certain national group with its own traditions and cultural background and its own peculiar complex of stimulating conditions, he will display the behavioral characteristics of that group regardless of his racial origin. It should be understood, of course, that mere physical presence does not constitute group membership in a psychological sense. Thus if a Negro child were brought up in a community composed exclusively of whites, he would not necessarily receive the same social stimulation as a white child. Similarly, a boy who is brought up exclusively by female relatives will not develop the personality traits of a girl. A psychological group is based solely upon shared behavior and not upon geographical proximity or biological resemblance.

It follows from such a concept of group that any one individual is effectively a member of a large and varied set of groups. A *multiplicity of behavioral groups*, large and small, cut across each other in the individual's background. Some of the most important of these groups have already been discussed in Part III of the present book. The individual is born into a broad cultural division such as, for example, "Western civilization," with its characteristic sources of stimulation. He will develop certain aptitudes, emotional traits, attitudes, and beliefs as a result of his affiliation with this group. He is also a member of a given national group with its more specific traditional ways of acting.

[7] This criterion of a psychological group is essentially that formulated by Kantor (23), who seems to have been the first to discuss social behavior in terms of *shared responses* to objects having common stimulus functions.

If the individual displays certain physical characteristics, such as a particular skin color, facial conformation, and body build, he may be classified as a member of a given "racial" group which occupies a distinct position within the broader national division. In so far as his racial background leads to certain social distinctions and culturally imposed differentiations of behavior, it will operate as an effective grouping. The same may be said of sex. If, within a given society, traditional beliefs in regard to sex differences exist so that the sexes are exposed to dissimilar psychological stimulation, then the individual's sex will in part determine his behavioral characteristics.

There are a number of other behavioral groupings which, although less frequently recognized and less clearly defined, may be equally influential in the individual's development. Thus it will be recalled that important psychological differences are usually found between the city-bred and country-bred child, as well as between different social-status classes (cf. Ch. 23). Similarly, the particular state, province, or other major division of a nation in which the individual is reared, and even the specific town and neighborhood in which he lives, will exert significant influences upon his intellectual and emotional development.

Other groups with which an individual identifies himself behaviorally are his occupational class, his religious sect, his political party, his club, his educational institution. That such groupings represent clear-cut cultural distinctions is readily illustrated by the stereotypes which have become attached to many of these groups. To people within our society, a distinct picture will be suggested by the mention of such designations as country doctor, business man, Roman Catholic, Orthodox Jew, Republican, Rotarian, Harvard man. These groups influence the individual's behavior in two ways. First, they directly stimulate and foster certain ways of acting. Secondly, the reactions of other people to the individual are influenced by their knowledge of his group affiliation. The social attitudes and "social expectancy" which the individual encounters will in turn affect his behavior.

Family groupings, with their characteristic activities and traditions, constitute another important part of the individual's psychological environment. The famous Herreshoff family of boat-designers and builders, the degenerate Kallikaks, eminent families such as the Huxleys and the Darwins, and many other striking examples testify to the cultural influence of family membership. Cutting across such

family groupings are age distinctions. "Stages" are socially imposed upon the continuous life activities of the individual and he is treated more or less differently at each period. The individual may also look upon himself as belonging to a particular generation—he may be a member of the "older generation," the "young married set," the "teen-agers," and so forth. Even such apparently minor factors as one's hobbies and recreations will in turn affect the individual's subsequent behavior. Psychological membership in many new groups may result from a newly developed interest in bowling, stamp collecting, or early American pressed glass. The number of behavioral groupings could easily be multiplied. These examples will suffice to illustrate the nature of such groupings and their effect upon the individual.

THE MEANING OF INDIVIDUALITY

The individual may be regarded partly as a resultant of his multiple group memberships. To be sure, each individual also undergoes experiences which are absolutely unique to himself. Such experiences are probably less significant, however, in shaping the more basic aspects of his personality than is his shared behavior. The experiences which are common to a group of individuals have a certain degree of permanence in the sense that they will tend to be repeated more often and to be corroborated or reenforced by other similar experiences. In general, the more highly organized the group, the more consistent and systematic will be the experiences which its members undergo. This will tend to make the shared experiences on the whole more effective than the purely individual. Moreover, even the individual's idiosyncratic experiences will generally have certain cultural features which differentiate them from the idiosyncratic experiences of persons in other cultures. Thus an individual may compose a poem which is unique in its totality and to this extent unlike any poem ever written by any other person; but the fact that the poem is a political satire, that it is composed in the English language and in iambic pentameter, and that it is written with a ball-point pen are among the many distinctly cultural features of such an activity.

In view of the pronounced effect of such shared or common behavior upon the individual's development, it may appear surprising that individuals are no more alike in their behavior repertoire than we ordinarily find them to be. The extent of individual differences

within any one group is extremely large. In fact, the variations among individuals have always proved to be more marked than the differences from one group to another. How can the "individuality" of each person be explained in terms of his shared experiential background?

The key to this problem seems to lie in the *multiplicity* of overlapping groups with which the individual may be behaviorally identified. The number of such groups is so great that the *specific combination* is unique for each individual. Not only does this furnish a stimulational basis for the existence of wide individual differences, but it also suggests a mechanism whereby the individual may "rise above" his group. There are many examples of individuals who have broken away from the customs and traditional ways of acting of their group. Through such situations, modifications of the group itself may also be effected.

In these cases the individual is not reacting contrary to his past experience, as might at first appear. This would be psychologically impossible. His behavior is the result of psychological membership in various *conflicting* groups. Many group memberships can exist side by side in a composite behavioral adjustment. But in certain cases two or more groups may foster different ways of reacting to the same situation. This enables the individual to become aware of the arbitrariness of the restrictions and traditions of each group, to evaluate them critically, and to regard them more "objectively." Membership in many diverse groups frees the individual from the intellectual and other limitations of each group and makes possible the fullest development of "individuality."

Author Index

Subject Index